FILE COPY

AD-A221 466

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
February, 1990

3. REPORT TYPE AND DATES COVERED
Technical Report , Number 90-247

4. TITLE AND SUBTITLE
A HYPOTHESIS TEST OF CUMULATIVE SUMS OF MULTINOMIAL PARAMETERS[1]

5. FUNDING NUMBERS
AFOSR-88-~~0040~~ 0040
61102F 2304/A5

6. AUTHOR(S)
J. H. Clair[2] and D. A. Meeter

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Statistics
Florida State University
Tallahassee, FL 32306-3033

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR- 90-0426

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NM
Bldg. 410
Bolling AFB, DC 20332-6448

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFOSR-88-0040

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
SAME AS PAGE ONE OF REPORT.

Accession For	
NTIS GRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

DTIC
ELECTE
MAY 1 1990
S E D

14. SUBJECT TERMS

15. NUMBER OF PAGES
37

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
UNCLASSIFIED

18. SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

19. SECURITY CLASSIFICATION OF ABSTRACT
UNCLASSIFIED

20. LIMITATION OF ABSTRACT
SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

AFOSR-TR- 90-0420

A Hypothesis Test of Cumulative Sums of Multinomial Parameters[1]

by

J. H. Clair[2] and D. A. Meeter

FSU Technical Report Number M-821

AFOSR Technical Report Number 90-247

Department of Statistics
Florida State University
Tallahassee, FL 32306-3033

February, 1990

[1]Research partly supported by the Air Force Office of Scientific Research under Contract Number AFOSR 88-~~0400~~. 0040

[2]Currently at Merck Sharpe & Dohme Research Labs.

A Hypothesis Test of Cumulative Sums of Multinomial Parameters[1]

J. H. Clair[2] and D. A. Meeter
Florida State University

Abstract for Air Force Accomplishments

Mr. Jerry Klion, Griffiss Air Force Base, Rome, New York, posed the following problem: The Air Force is considering the contract renewal application of a civilian contractor hired to maintain in working order a series of radar stations. The measure of performance of interest is T, the time that a particular station is not 'on line' while being down for repair. The contract stipulates that repair service will be such that on the average 50% of all repairs will be completed before L_1 hours and 90% of all repairs shall be completed before L_2 hours. It also states that the repair contract will be renewed on the basis of a decision rule that errs by failing to renew when the case is that the contract should be renewed with a probability of α. The renewal of the contract depends on the making a decision based on N repair times, $T_1, T_2, \cdots, T_N$, of the contractor as to whether or not L_1 is *at least* the 50^{th} percentile and L_2 is *at least* the 90^{th} percentile of $F(\cdot)$, the distribution function F(.) of these repair times. The usual test based on the binomial distributions of the number of repairs before L_1 and the number of repairs before L_2 suffers from two problems: (1) Its true size is at times far from the nominal size and (2) because of the discrete of the random variables, cannot be performed at the stipulated size. This paper proposes the use of the likelihood ratio test based on the nultinomial joint distribution of the number of repairs before L_1 and the number of repairs before L_2.

More generally, a likelihood ratio test to simultaneously test K hypotheses on cumulative sums of multinomial parameters is described here. An algorithm is supplied for the evaluation of this statistic. It is shown that this test procedure allows for a test at approximately the correct size that has uniformly better power with respect to test that is the natural competitor.

Key Words: hypothesis testing, multinomial, likelihood ratio, quantiles

The Problem and its Notation

The following problem was posed: one of the armed forces is considering the contract renewal application of a civilian contractor hired to maintain in working order a series of radar stations. The measure of performance of interest is T, the time that a particular station is not 'on line' while being down for repair. The contract stipulates that repair service will be such that 50% of all repairs will be completed before L_1 hours and 90%

[1]Research partially sponsored by the U.S. Air Force Office of Scientific Research under AFOSR Grant 88-0400.

[2]Currently at Merck Sharpe & Dohme Research Labs.

of all repairs shall be completed before L_2 hours. It also states that the repair contract will be renewed on the basis of a decision rule that errs by failing to renew when the case is that the contract should be renewed with a probability of α. The renewal of the contract depends on the making a decision based on N repair times, $T_1, T_2, \cdots, T_N$, of the contractor as to whether or not L_1 is *at least* the 50^{th} percentile and L_2 is *at least* the 90^{th} percentile of $F(\cdot)$, the distribution function of these repair times. In this situation, N the number of repair times may be small, usually less than 20.

In more general terms, a contract might stipulate performance criteria on K quantiles $(K \geq 2)$ of a distribution function $F(\cdot)$ specified by a set of time periods, $\{L_1, L_2, \cdots, L_K\}$, $L_1 < L_2 < \cdots < L_K$, and a set of probabilities, $\{p_{0\,1}, p_{0\,2}, \cdots, p_{0\,K}\}$, $0 \leq \sum_{\nu=1}^{i} p_{0\,\nu} \leq 1$ for $1 \leq i \leq K$. To simplify the notation, let $p_i = F(L_i) - F(L_{i-1})$ and

$$P_i^j = \sum_{\nu=i+1}^{j} p_\nu \text{ and } {P_0}_i^j = \sum_{\nu=i+1}^{j} p_{0\,\nu} \text{ for } i \leq j.$$

When the subscript i is 0, it will be suppressed:

$$P^j \equiv P_0^j \text{ and } {P_0}^j \equiv {P_0}_0^j.$$

On the basis of N observations, the interest is in simultaneously testing the following K hypotheses with a specified type I error rate α:

$$\mathrm{H}_0^i: \quad P^i \geq {P_0}^i, \; i = 1, 2, \cdots, K. \tag{1}$$

The general problem is to simultaneously test K one-sided hypotheses on cumulative sums of parameters of a multinomial distribution. Although the discussion here focuses on K hypotheses stipulating a greater than or equal $(\geq)$ relationship, the techniques illustrated here can also be applied to hypotheses in the opposite direction $(\leq)$.

The Standard Approach and its Limitations

For $\nu = 1, \cdots, K$, let $N_\nu =$ the number of repair times that fall between L_ν and $L_{\nu-1}$ with $L_0 \equiv 0$. More generally N_ν is the number of observations falling in the $\nu^t h$ cell of a multinomial distribution with parameter p_ν.. The quantity $\sum_{\nu=1}^{i} N_\nu$ is a binomial random variable with parameter P^i, $i = 1, \cdots, K$. A reasonable approach for constructing a test of the hypothesis in (1) is employ the **Union-Intersection** technique : perform K tests on the P^i, $i = 1, 2, \cdots, K$. The null hypothesis is rejected if for any i, $1 \leq i \leq K$, $\sum_{\nu=1}^{i} N_\nu < n_i^0$. The critical values n_i^0 are either determined from the binomial distribution or from the normal approximation to the binomial such that for $i = 1, \cdots, K$,

$$\max_{P^i \geq {P_0}^i} \left\{ P\left(\sum_{\nu=1}^{i} N_\nu < n_i^0 \right) \right\} = P_{{P_0}^i}\left(\sum_{\nu=1}^{i} N_\nu < n_i^0 \right) = \alpha_i.$$

Here α_i is the error rate or size for testing the individual hypothesis H_0^i. The n_i^0 is be selected in such a way to make the overall experiment-wise error rate conform to

the contract- specified error level α . This choice of the n_i^0 can utilize the Bonferroni Inequality:

$$P\left\{\bigcup_{i=1}^{K}\left\{\sum_{\nu=1}^{i} N_\nu < n_i^0\right\}\right\} \leq \sum_{i=1}^{K} P\left\{\sum_{\nu=1}^{i} N_i < n_i^0\right\} = \sum_{i=1}^{K} \alpha_i. \tag{2}$$

When the null hypothesis is true and $\sum_{i=1}^{K} \alpha_i = \alpha$, the specified experiment-wise error rate is bounded above by α.

In situations in which we are testing a hypothesis on $K \geq 3$ quantiles, for small samples the union-intersection procedure gives the decision maker little guidance in either constructing a critical region or in estimating the error level. For example, below is a portion of the binomial tables for $N = 8$.

Table 1: Abbreviated Binomial Table for $N = 8$.

n \ p	.20	.40	.60	.80
0	.167	.017	.001	.000
1	.503	.106	.009	.000
2	.797	.315	.050	.001
3	.944	.594	.174	.010
4	.990	.826	.406	.056
5	.999	.950	.685	.203
6	1.000	.992	.894	.497
7	1.000	.999	.983	.832
8	1.000	1.000	1.000	1.000

In testing the hypothesis

$$P_0^1 \geq .20, \;\; P_0^2 \geq .40, \;\; P_0^3 \geq .60, \;\; \text{and } P_0^4 \geq .80 \text{ at } \alpha = .05,$$

the cautious decision-maker, using the Union-Intersection technique, would feel obliged to reject if $\sum_{\nu=1}^{i} N_\nu = 0$ for $i = 1, 2, 3$, or 4, knowing only that the type I error rate is bounded above by $.167 + .017 + .001 = .185$. A selection of any other critical region would more than likely rule out any test at all on the 20^{th} percentile and call for a test based only on (n_2, n_3, n_4). This reduces to a test on three proportions. But we wanted to test four! One concludes that for small samples, where the probability of 1 'success' at P^1 is greater than the required size of the test, the union-intersection technique is inadequate.

A Generalized Likelihood Ratio Test (GLRT)

From (1) we note that the hypothesis involves a test of multinomial parameters. The observed vector $(N_1, N_2, \cdots, N_K)$ is distributed as a K-nomial random vector with

parameter $(p_1, p_2, \cdots, p_K)$ in the parameter space

$$\Omega = \left\{ \vec{p} \in R^K \mid 0 \le \sum_{\nu=1}^{K} p_\nu \le 1,\ p_i \ge 0,\ 1 \le i \le K \right\}.$$

The null hypothesis tests if $\vec{p} \in \Omega_0$ with

$$\Omega_0 = \left\{ \vec{p} \in \Omega \mid P^i \ge P_0{}^i,\ 1 \le i \le K \right\}.$$

The vector $(p_{o1}, p_{o2}, \cdots, p_{oK})$ characterizes the population with respect to the K disjoint time intervals $L_i - L_{i-1}$, $i = 1, \cdots, K$, in the sense that $P_0{}^i$ is the lower bound for the hypothesized proportion of repair times occurring before L_i.

An example for $K = 2$ is illustrative.

<u>EXAMPLE 1</u>.

The parameter space is

$$\Omega = \left\{ (p_1, p_2) \mid 0 \le p_1 \le 1,\ 0 \le p_2 \le 1,\ P^2 \le 1 \right\}.$$

Partition Ω into

$$\Omega_0 = \left\{ (p_1, p_2) \mid p_1 \ge p_{o1},\ P^2 \ge P_0{}^2 \right\} \text{ and } \Omega_0^c = \Omega - \Omega_0.$$

The null hypothesis space, Ω_0, is shaded in Figure 1.

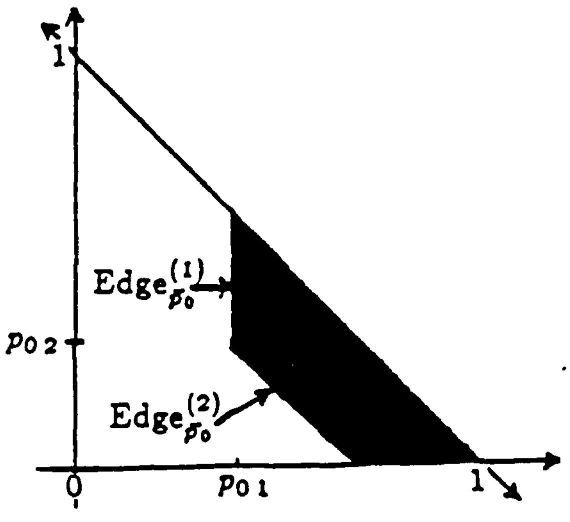


Figure 1 Partitioning the Parameter Space when K = 2

We wish to test

$$\mathrm{H}_0 : \ (p_1, p_2) \in \Omega_0 \text{ versus } \mathrm{H}_1 : \ (\dot{p_1}, p_2) \in \Omega_0^c.$$

A natural suggestion for testing this hypothesis is to employ a generalized likelihood ratio test (GLRT). Many statistical tests are based on the generalized likelihood ratio approach and there is a large body of statistical theory in support of such procedures. The large sample theory for the particular application here was first presented by Chernoff (1954). Extensions to the Chernoff result can be found in Feder (1968) ,and Self and Liang (1987). A recent application can be found in Greenberg (1985). While Feder and Self and Liang address general properties of this statistic in the large sample case, Greenberg discusses a specific application similar to the problem addressed here. Greenberg's interest is to simultaneously test one- sided hypotheses on $K-1$ multinomial parameters from a K-nomial distribution. In terms of (1) Greenberg wishes to test

$$\mathrm{H}_0^i : \quad P^i \geq P_0{}^i, \ i = 2, \cdots, K. \tag{1'}$$

Although the difference between (1) and (1') is only the lack of interest in what happens in the first cell, the solution to the maximization problem is made more tractable as Greenberg shows. Here we consider the larger hypothesis (1) and construct the test procedure when applied to small sample case ($N \leq 20$).

As before, N denotes the total number observations and let

$$\vec{N} = \{N_1, N_2, \cdots N_K\}.$$

The generalized likelihood ratio used here is defined to be

$$\lambda(N) = \frac{\max_{\vec{p} \in \Omega_0} L(\vec{N}, \vec{p})}{\max_{\vec{p} \in \Omega_0^c} L(\vec{N}, \vec{p})} \text{ where } L(\vec{N}, \vec{p}) = \prod_{\nu=1}^{K} p_\nu^{N_\nu} \left(1 - P^K\right)^{\left(N - \sum_{\nu=1}^{K} N_\nu\right)}.$$

Denote

$$\frac{1}{N} \log(\lambda(N)) \equiv \log L\left(\hat{\vec{p}}, \vec{p}\right) \equiv \sum_{\nu=1}^{K} \frac{N_\nu}{N} \log(p_\nu) + \left(1 - \sum_{\nu=1}^{K} \frac{N_\nu}{N}\right) \log\left(1 - P^K\right).$$

The decision rule is

$$\text{reject } \mathrm{H}_0 \text{ if } \lambda(N) < \delta \text{ where } \max_{\vec{p} \in \Omega_0} \mathrm{P}\left\{\lambda(N) < \delta\right\} = \alpha.$$

<u>Evaluating $\lambda(N)$</u>

The use if the Likelihood Ratio statistic requires the determination of δ based on the distribution of $\lambda(N)$. The evaluation of the statistic involves maximizing $\log L(\hat{\vec{p}}, \cdot)$ over

Ω_0 and its complement. The geometric structure of the null hypothesis space and algebraic properties of the likelihood function in this case make it possible to establish an algorithm that facilitates the evaluation process. The formal proofs are detailed and reserved for the Appendix. However, intuitive arguments and examples are given to help the reader understand better the rationale for the algorithm and the manner in which it functions.

Assume in what follows that $\Omega \subset R^K$. Then following definitions and results hold.

(a) If $\hat{\vec{p}}$ (the global MLE) $\in \Omega_0^c(\Omega_0)$ then $L(\vec{n},\vec{p})$ is maximized on $\Omega_0(\Omega_0^c)$ on the boundary between Ω_0 and $\Omega_0^c \equiv \mathbf{B}_{\vec{p}_0}$. It is well known that the multinomial pdf converges to the multivariate normal pdf implying that the log multinomial likelihood function converges to the the log of normal likelihood function which in turn is inversely proportional to a (Mahalanobis) distance. Thus, given an observation $\hat{\vec{p}}$, $\log L(\hat{\vec{p}}, \vec{p})$ behaves like aa inverted metric: if $\hat{\vec{p}} \in \Omega_0^c(\Omega_0)$ and $\vec{p}^* \in \Omega_0(\Omega_0^c)$, then $\log L(\hat{\vec{p}}, \vec{p})$ decreases as a function of $\vec{p}$ along the line segment $\mathcal{L}$ between $\hat{\vec{p}}$ and $\vec{p}^*$. Because $\mathcal{L}$ must cross $\mathbf{B}_{\vec{p}_0}$ before it enters $\Omega_0(\Omega_0^c)$, it follows that $\log L(\hat{\vec{n}}, \vec{p})$ is maximized for $\vec{p} \in \Omega_0(\Omega_0^c)$ on $\mathbf{B}_{\vec{p}_0}$.

(b) Because of (a), the evaluation of $\lambda(N)$ depends on points in $\mathbf{B}_{\vec{p}_0}$. Here we define the constituent parts of $\mathbf{B}_{\vec{p}_0}$:

Define a "one-restriction" plane as

$$\text{Plane}_{\vec{p}_0}^{(i)} \equiv \left\{\vec{p} \in R^K \,\middle|\, P^i = {P_0}^i\right\}.$$

In turn a "t-restriction" plane is defined to be for $2 \leq t \leq K$ and any nonempty subset $\{i_1, \cdots, i_t\}$ of the set $\{1, 2, \cdots, K\}$ as

$$\text{Plane}_{\vec{p}_0}^{(i_1,\cdots,i_t)} \equiv \bigcap_{\nu\in\{i_1,\cdots,i_t\}} \text{Plane}_{\vec{p}_0}^{(\nu)},$$

Define

$$\text{Edge}_{\vec{p}_0}^{(i_1,\cdots,i_t)} = \text{Plane}_{\vec{p}_0}^{(i_1,\cdots,i_t)} \cap \Omega_0 \equiv$$

$$\left\{\vec{p} \in R^K \cap \Omega \,\middle|\, P^i = {P_0}^i \quad \text{for } i = i_1, i_2, \cdots, i_t, P^i > {P_0}^i \text{ for } i \neq i_1, i_2, \cdots, i_t\right\}.$$

The boundary, $\mathbf{B}_{\vec{p}_0}$, is the nondisjoint union of the $2^K - 1 = \sum_{\nu=1}^{K} \binom{K}{\nu}$ edges.

For example, if $K = 3$, there are $2^3 - 1 = 7$ such edges whose union determines $\mathbf{B}_{\vec{p}_0}$. If ${P_0}^1 = .2, {P_0}^2 = .4$, and ${P_0}^3 = .6$, then one of the $\binom{3}{2}$ '2-restriction' planes is based on the subset $\{1, 2, \}$ of the set $\{1, 2, 3\}$ denoted above as

$$\text{Plane}_{\vec{p}_0}^{(1,2)} \equiv \bigcap_{\nu\in\{1,2\}} \text{Plane}_{\vec{p}_0}^{(\nu)} \equiv \left\{\vec{p} \in R^3 \cap \Omega \,\middle|\, P^1 = .2, P^2 = .4\right\}$$

and

$$\text{Edge}_{\vec{p}_0}^{(1,2)} \equiv \left\{\vec{p} \in R^3 \cap \Omega \,\middle|\, P^1 = .2, P^2 = .4, P^3 > .6\right\}.$$

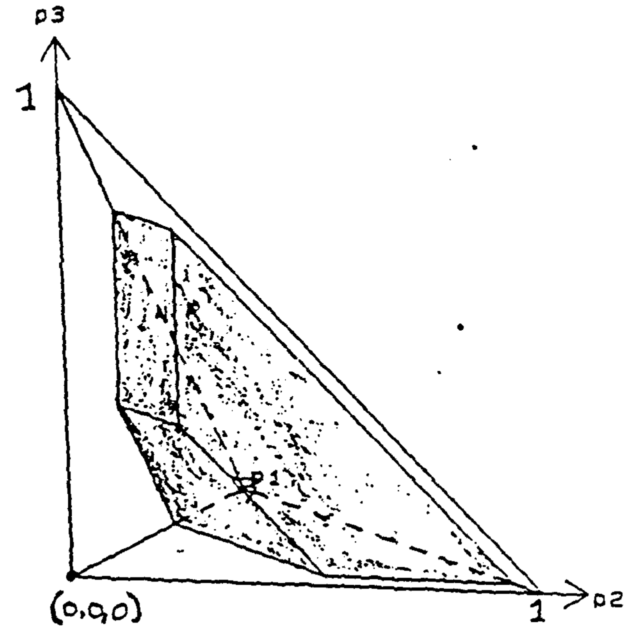


Figure 2(a) For $K = 3$, Ω with Ω_0, the boundary $B_{\bar{p}_0}$ is shaded.

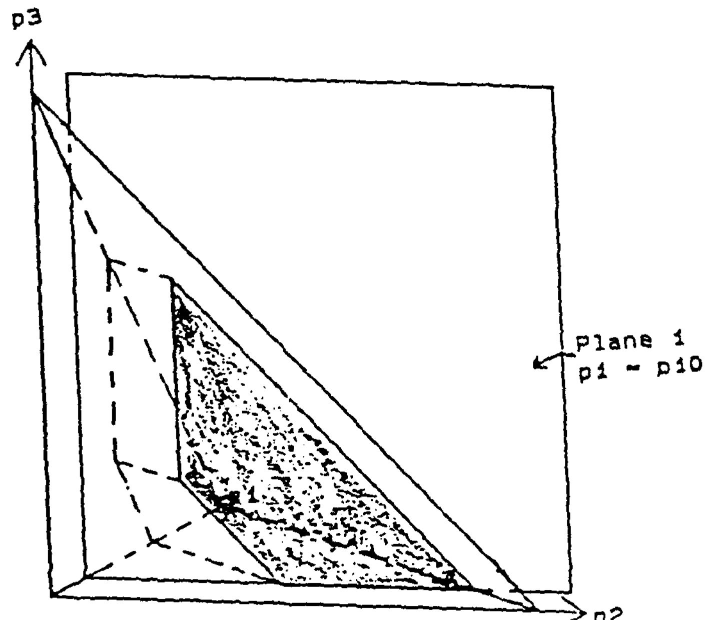


Figure 2 (b) Ω and $\text{Plane}^{(1)}_{\bar{p}_0}$, $\text{Plane}^{(1)}_{\bar{p}_0} \cap B_{\bar{p}_0}$ is shaded.

Figure 2 (a) illustrates Ω for this example. There the boundary $B_{\bar{p}_0}$ is shaded. The parameter space Ω is the set of points inside the tetrahedron Ω. The null hypothesis space Ω_0 is all of the points on $B_{\bar{p}_0}$ and 'behind' it. Figures 2 (b) , (c), and (d) show extended rectangular portions of $\text{Plane}^{(i)}_{\bar{p}_0}$ for $i = 1, 2$, an 3 respectively, indicating the contribution of each plane to $B_{\bar{p}_0}$.

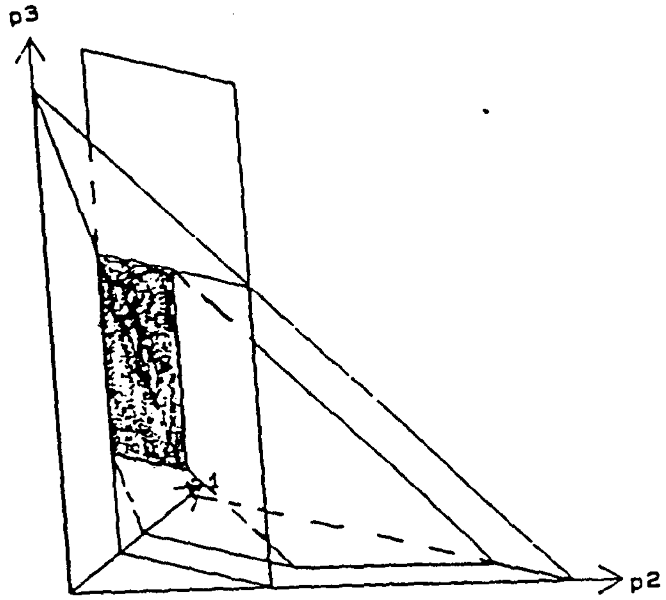


Figure 2 (c) Ω and $\text{Plane}_{\bar{p}_0}^{(2)}$, $\text{Plane}_{\bar{p}_0}^{(2)} \cap B_{\bar{p}_0}$ is shaded.

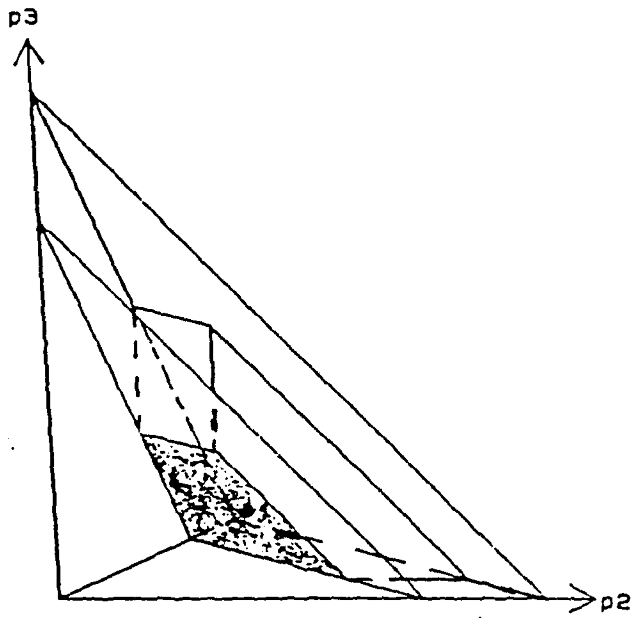


Figure 2 (d) Ω and $\text{Plane}_{\bar{p}_0}^{(3)}$, $\text{Plane}_{\bar{p}_0}^{(3)} \cap B_{\bar{p}_0}$ is shaded.

For $\hat{\vec{p}}$, define $\hat{p}^{(i_1,\cdots,i_t)}$ as the vector that maximizes $\log L(\hat{\vec{p}},\vec{p})$ on $\text{Plane}_{\vec{p}_0}^{(i_1,\cdots,i_t)}$. Consistent with the earlier shorthand notation for the cumulative sums, we denote

$$\hat{P}^{(i_1,\cdots,i_t)}{}_i^j \equiv \sum_{\nu=i+1}^{j} \hat{p}_\nu^{(i_1,\cdots,i_t)} \quad \text{and } \hat{P}^{(i_1,\cdots,i_t)j} \equiv \hat{P}^{(i_1,\cdots,i_t)}{}_0^j.$$

(c) For an observed $\hat{\vec{p}}$ the coordinates of $\hat{p}^{(i_1,\cdots,i_t)}$ are given by the following theorem.

THEOREM 1.

If we denote 0 as i_0, then the i^{th} coordinate of $\hat{\vec{p}}^{(i_1,\cdots,i_t)}$ is

$$\hat{p}_i^{(i_1,\cdots,i_t)} = \begin{cases} (a)\dfrac{\hat{p}_i\, P_0{}_{i_{j-1}}^{i_j}}{\hat{P}_{i_{j-1}}^{i_j}}, & \text{if } i_{j-1} < i \le i_j,\ 1 \le j \le t, \text{ and } \hat{P}_{i_{j-1}}^{i_j} \neq 0, \\ (b)\quad p_{0\,i}, & \text{if } i_{j-1} < i \le i_j,\ 1 \le j \le t, \text{ and } \hat{P}_{i_{j-1}}^{i_j} = 0, \\ (c)\quad \dfrac{\hat{p}_i\left(1 - P_0{}^{i_t}\right)}{1 - \hat{P}^{i_t}}, & \text{if } i_t < i \le K \text{ and } \hat{P}^{i_t} \neq 1, \\ (d)\quad 0, & \text{if } i_t \le i \le K \text{ and } \hat{P}^{i_t} = 1. \end{cases}$$

We illustrate the use of Theorem 1 and its notation with a few examples.

EXAMPLE 2.

Suppose that $K = 4$. Assume that $\hat{p}_i \neq 0,\ 1 \le i \le K$.

(1). We describe, for a particular $\hat{\vec{p}}$, the coordinates of $\hat{\vec{p}}^{(1)}$, the point that maximizes $\log L(\hat{\vec{p}},\vec{p})$ on the one-restriction plane

$$\text{Plane}_{\vec{p}_0}^{(1)} = \{\vec{p} \in \Omega \mid p_1 = p_{0\,1}\}.$$

Here t, the number of restrictions is 1, $i_t = i_1 = 1$, and $0 < j \le i_t = 1$. Thus j only takes the value 1. The subscript i takes the values 1 to 4.
For $i = 1$, we use formula part (a)

$$\hat{p}_1^{(1)} = \frac{\hat{p}_1\left(P_0{}^1\right)}{\hat{P}^1} = \frac{\hat{p}_1 p_{0\,1}}{\hat{p}_1} = p_{0\,1}.$$

Note that if $\hat{p}_1 = 0$ we would use formula part (b) and simply assign $\hat{p}_1^{(1)} = p_{0\,1}$ since $\hat{p}_1^{(1)}$ must be an element of $\text{Plane}_{\vec{p}_0}^{(1)}$.

For $i_1 < i \leq 4$ or for $2 \leq i \leq 4$, we use formula part (c) to compute

$$\hat{p}_i^{(1)} = \frac{\hat{p}_i\left(1 - P_0{}^1\right)}{1 - \hat{P}^1}.$$

Thus

$$\hat{p}_2^{(1)} = \frac{\hat{p}_2(1 - p_{0\,1})}{1 - \hat{p}_1},$$
$$\hat{p}_3^{(1)} = \frac{\hat{p}_3(1 - p_{0\,1})}{1 - \hat{p}_1}, \text{ and}$$
$$\hat{p}_4^{(1)} = \frac{\hat{p}_4(1 - p_{0\,1})}{1 - \hat{p}_1}.$$

Note that if any of the $\hat{p}_i = 0$ for i = 2, 3, or 4, then naturally the value of $\hat{p}_i^{(1)}$ would be 0.

If $\hat{P}^1 = \hat{p}_1 = 1$, then the likelihood function depends only on the value of p_1. From formula part (d) we assign $\hat{p}_i^1$ the value 0 for $2 \leq i \leq 4$.

(2). We compute the coordinates of $\hat{\vec{p}}^{(3)}$, the point that maximizes $\log L(\hat{\vec{p}}, \vec{p})$ on the one-restriction plane

$$\text{Plane}_{\vec{p}_0}^{(3)} = \left\{\vec{p} \in \Omega \,\middle|\, P^3 = P_0{}^3\right\}.$$

In this case also $t = 1$ and $i_t = 3$. Again j takes the value 1. For $1 \leq i \leq i_j = i_1 = 3$. If $\hat{P}^3 \neq 0$, we use formula part (a):

$$\hat{p}_i^{(3)} = \frac{\hat{p}_i\left(P_0{}^3\right)}{\hat{P}^3} = \frac{\hat{p}_i(p_{0\,1} + p_{0\,2} + p_{0\,3})}{\hat{p}_1 + \hat{p}_2 + \hat{p}_3}.$$

The likelihood function restricted this set is

$$\sum_{\nu=1}^{2} \hat{p}_\nu \log(p_\nu) + \hat{p}_3 \log\left(P_0{}^3 - p_1 - p_2\right) + \left(1 - \hat{P}^4\right) \log\left(1 - P_0{}^3 - p_4\right).$$

Note that if $\hat{P}^3 = 0$, the likelihood function does not depend on p_1, p_2, or p_3. However we require that $\hat{P}^{(3)3} = P_0{}^3$. Since the values p_1, p_2, and p_3 have no effect on the value of the likelihood function, we assign to them the values $p_{0\,1}$, $p_{0\,2}$, and $p_{0\,3}$ respectively. This is an example of the application of formula part (b).
For $i = 4$, we use formula part (c):

$$\hat{p}_4^{(3)} = \frac{\hat{p}_4\left(1 - P_0{}^3\right)}{1 - \hat{P}^3} = \frac{\hat{p}_4(1 - p_{0\,1} - p_{0\,2} - p_{0\,3})}{1 - \hat{p}_1 - \hat{p}_2 - \hat{p}_3}.$$

(3). Now we compute $\hat{\vec{p}}^{(1,3)}$, the point that maximizes $\log L(\hat{\vec{p}}, \vec{p})$ on the two-restriction plane

$$\text{Plane}_{\vec{p}_0}^{(1,3)} \equiv \left\{\vec{p} \in \Omega \,\middle|\, p_1 = p_{0\,1} \text{ and } P^3 = P_0{}^3\right\}.$$

Here $t = 2$, $i_1 = 1$, and $i_2 = 3$. For $i = 1$, using formula part (a)

$$\hat{p}_1^{(1,3)} = \frac{\hat{p}_1\left(P_0{}^1\right)}{\hat{P}^1} = \frac{\hat{p}_1 p_{0\,1}}{\hat{p}_1} = p_{0\,1}.$$

For $i = 2$ and 3,

$$\hat{p}_2^{(1,3)} = \frac{\hat{p}_2 P_0{}_2^3}{\hat{P}_2^3} = \frac{\hat{p}_2(p_{0\,2} + p_{0\,3})}{\hat{p}_2 + \hat{p}_3} \text{ and}$$

$$\hat{p}_3^{(1,3)} = \frac{\hat{p}_3 P_0{}_2^3}{\hat{P}_2^3} = \frac{\hat{p}_3(p_{0\,2} + p_{0\,3})}{\hat{p}_2 + \hat{p}_3}.$$

For $i = 4$, we use formula part (c):

$$\hat{p}_4^{(1,3)} = \frac{\hat{p}_4\left(1 - P_0{}^3\right)}{1 - \hat{P}^3} = \frac{\hat{p}_4(1 - p_{0\,1} - p_{0\,2} - p_{0\,3})}{1 - \hat{p}_1 - \hat{p}_2 - \hat{p}_3}.$$

Note that $\hat{p}_1^{(1,3)} = \hat{p}_1^{(1)}$ and $\hat{p}_4^{(1,3)} = \hat{p}_4^{(3)}$.

(4). Now we will compute $\hat{\vec{p}}^{(2,4)}$, the point that maximizes $\log L(\hat{\vec{p}}, \vec{p})$ on

$$\text{Plane}_{\vec{p}_0}^{(2,4)} = \left\{\vec{p} \in \Omega \mid P^2 = P_0{}^2 \text{ and } P^4 = P_0{}^4\right\}.$$

Again $t = 2$, $i_1 = 2$, and $i_2 = 4$. For $i = 1$ and 2,

$$\hat{p}_i^{(2,4)} = \frac{\hat{p}_i\left(P_0{}^2\right)}{\hat{P}^2} = \frac{\hat{p}_i(p_{0\,1} + p_{0\,2})}{\hat{p}_1 + \hat{p}_2}.$$

For $i = 3$ and 4,

$$\hat{p}_i^{(2,4)} = \frac{\hat{p}_i\left(P_0{}_3^4\right)}{\hat{P}_3^4} = \frac{\hat{p}_i(p_{0\,3} + p_{0\,4})}{\hat{p}_3 + \hat{p}_4}.$$

So far we have described the candidate set, $\mathrm{B}_{\vec{p}_0}$, where $\log L(\hat{\vec{p}}, \vec{p})$ should be maximized, identified sets denoted by $\text{Plane}_{\vec{p}_0}^{(i_1, i_2, \cdots, i_t)}$, subsets of which are the constituent parts of $\mathrm{B}_{\vec{p}_0}$, and determined the components of $\hat{\vec{p}}^{(i_1, \cdots, i_t)}$ that maximize $\log L(\hat{\vec{p}}, \vec{p})$ on each $\text{Plane}_{\vec{0}}^{(i_1, i_2, \cdots, i_t)}$. Thus for an observed $\hat{\vec{p}} \in \Omega_0^c$, there are $2^K - 1$ candidate $\hat{\vec{p}}^{(i_1, \cdots, i_t)}$'s for $\hat{p}_0$, the $\vec{p}$ that maximizes $\log L(\hat{\vec{p}}, \vec{p})$ on Ω_0 when $\hat{\vec{p}} \in \Omega_0^c$. Now we give the criteria to chose $\hat{\vec{p}}_0$ from among these $2^K - 1$ candidates.

(d) If $\hat{p} \in \Omega_0^c$ then $\hat{\vec{p}}_0 = \hat{\vec{p}}^{(i_1, \cdots, i_t)}$, where $\{i_1, \cdots, i_t\}$ is the subset of $\{1, 2, \cdots, K\}$ with the smallest number of elements such that $\hat{\vec{p}}^{(i_1, \cdots, i_t)} \in \Omega_0$.

Intuitively, if A and B are sets with $A \subset B \subset \{1, 2, \cdots, K\}$ then $\log L(\hat{\vec{p}}, \vec{p})^{(B)} \leq \log L(\hat{\vec{p}}, \vec{p}^{(A)})$. Thus parts of $\mathrm{B}_{\vec{p}_0}$, characterized by more restrictions $i_1, i_2, \cdots i_t$ should yield smaller values of $\log L(\hat{\vec{p}}, \vec{p})$ than parts characterized by fewer restrictions. However,

since set inclusion is only a partial ordering, some care must be taken to rule out candidate sets from different sets of orderings with the same number of restrictions. This fact makes an nontrivial task of proving that (d) above is true.

The result in (d) can be used to construct the following algorithm for evaluating $\hat{\vec{p}}_0$.

EVALUATION ALGORITHM.

If $\hat{\vec{p}}$ is in Ω_0^c, then $\log L(\hat{\vec{p}}, \vec{p})$ is maximized over Ω_0^c at $\hat{\vec{p}}$.

The maximum value on Ω_0 can be found in the following fashion;

(1) First compute the MLE's for the "one-restriction" planes $\hat{\vec{p}}^{(i)}$ for $1 \leq i \leq K$. If for any i, $\hat{\vec{p}}^{(i)} \in \Omega_0$, then $\log L(\hat{\vec{p}}, \vec{p})$, is maximized at that point. Stop.

(2) If none of the $\hat{\vec{p}}^{(i)} \in \Omega_0$, then begin computing the MLE's for the the "two-restriction" planes, $\hat{\vec{p}}^{(i_1,i_2)}$ for $i_1 = 1, \cdots K-1$, and $i_2 = i_1+1, \cdots, K$. If for any pair (i_1, i_2), $\hat{\vec{p}}^{(i_1,i_2)} \in \Omega_0$, then $\log L(\hat{\vec{p}}, \vec{p})$, is maximized at that point. Stop.

(3) If the restricted MLE's for the one-restriction and two- restriction planes are not in Ω_0, continue the procedure established in (1) and (2) for the triples that denote the MLE's for the "three- restriction" planes. If none of those are in Ω_0 go on the MLE's on the 'four-restriction' planes etc...

EXAMPLE 3.

Suppose that $K = 4$. We wish to test a null hypothesis about 4 proportions:

$$\mathrm{H}_0 : p_1 \geq .2,\ p_1 + p_2 \geq .4,\ p_1 + p_2 + p_3 \geq .6, \text{ and } p_1 + p_2 + p_3 + p_4 \geq .8.$$

Here $p_{10} = p_{20} = p_{30} = p_{40} = .2$. Let $N = 10$, $n_1 = 1$, $n_2 = 1$, $n_3 = 7$, and $n_4 = 1$, so that $\hat{p}_1 = \hat{p}_2 = \hat{p}_4 = .1$ and $\hat{p}_3 = .7$. Since $\hat{p}_1 < .2$, $\hat{\vec{p}} \notin \Omega_0$. Thus

$$\sup_{\vec{p} \in \Omega_0^c} \log L(\hat{\vec{p}}, \vec{p}) = \log L(\hat{\vec{p}}, \hat{\vec{p}}).$$

The Evaluation Algorithm can be used to compute $\sup_{\vec{p} \in \Omega_0} \log L(\hat{\vec{p}}, \vec{p})$. We begin by computing $\hat{\vec{p}}^{(1)}$. Using the results of Theorem 1 and following step one of the algorithm, $t = 1$ and $i_t = 1$;

$$\hat{p}_0^{(1)} = .2,$$
$$\hat{p}_2^{(1)} = \frac{.1(1-.2)}{1-.1} = .088\bar{8},$$
$$\hat{p}_3^{(1)} = \frac{.7(1-.2)}{1-.1} = .622\bar{2}, \text{ and}$$
$$\hat{p}_4^{(1)} = \frac{.1(1-.2)}{1-.1} = .088\bar{8}.$$

Because $.2 + .088\overline{8} < .4$, $\hat{\vec{p}}^{(1)} \notin \Omega_0$. We go on to compute $\hat{\vec{p}}^{(2)}$. In this case $t = 1$, $i_t = 2$,

$$\hat{p}_0^{(2)} = \frac{.1(.2 + .2)}{.1 + .1} = .2,$$
$$\hat{p}_2^{(2)} = \frac{.1(.2 + .2)}{.1 + .1} = .2,$$
$$\hat{p}_3^{(2)} = \frac{.7(1 - .2 - .2)}{1 - .1 - .1} = .525, \quad \text{and}$$
$$\hat{p}_4^{(2)} = \frac{.1(1 - .2 - .2)}{1 - .1 - .1} = .075.$$

Note that $\hat{\vec{p}}^{(2)} \in \Omega_0$. The algorithm says that we can stop the computation:

$$\sup_{\vec{p} \in \Omega_0} \log L(\hat{\vec{p}}, \vec{p}) = \log L(\hat{\vec{p}}, \hat{\vec{p}}^{(2)}).$$

No evaluation of $\log L(\hat{\vec{p}}, \vec{p})$ was necessary until we found the edge on which the maximum occurred and there was no computation involving the other $\binom{4}{1} + \binom{4}{2} + \binom{4}{3} + 1 - 2 = 13$ edges.

This algorithm speeds up the computation of $\log(\lambda(N))$ considerably over the more laborious procedure of finding all of the restricted MLE's that are in Ω_0 and comparing the values of $\log L(\hat{\vec{p}}, \vec{p})$ at each of the points to find the maximum value. Thus for small N the distribution of $\lambda(N)$ can be easily computed and a p-value assigned by computing probabilities of multinomial outcomes.

The Evaluation Algorithm applies to the maximization of the likelihood function when $\hat{\vec{p}} \in \Omega_0^c$. It can not be applied when $\hat{\vec{p}} \in \Omega_0$. Although when the likelihood function restricted to Ω_0^c is maximized on $\mathbf{B}_{\vec{p}_0}$, the rigorous argument used to establish the key property of the algorithm does not follow. Essentially this is because Ω_0^c is a union rather than an intersection of sets whose elements are described by inequalities involving the P^i and $P_0{}^i$.

When $\hat{p} \in \Omega_0$, instead of developing an algorithm for maximizing the likelihood function on Ω_0^c, we simplify the maximization problem by requiring that our test be a member of the class of all tests of size α that contain at least one element of Ω_0^c in their acceptance region. Since $\lambda(N) \geq 1$ for observed $\hat{\vec{p}} \in \Omega_0$ and $\lambda(N) \leq 1$ for observed $\hat{\vec{p}} \in \Omega_0^c$, this restriction implies that the acceptance region of the GLRT must contain Ω_0. This in turn implies that

$$\max_{\vec{p} \in \Omega_0} P_{\vec{p}} \left\{ \hat{\vec{p}} \in \Omega_0^c \right\} > \alpha.$$

Because of Proposition E.2.c in Marshall and Olkin(1979), noting that every vector $\vec{p} \in \Omega_0$ majorizes $\vec{p}_0$, we can simplify this requirement to

$$P_{\vec{p}_0} \left\{ \hat{\vec{p}} \in \Omega_0^c \right\} > \alpha. \tag{1}$$

This restriction is helpful in evaluating the likelihood ratio and does not limit significantly the situations to which the test can be applied. However, for an easy check to

see if a specific test situation meets the requirement (1) above, use the fact that for any $1 \leq i \leq K$,

$$P_{P_0{}^i}\left\{\hat{P}^i < P_0{}^i\right\} \leq P_{\vec{p}_0}\left\{\Omega_0^c\right\}.$$

Thus if for any i, $P_{P_0{}^i}\left\{\hat{P}^i < P_0{}^i\right\} \geq \alpha$, the testing situation meets the restriction.

This restriction makes sense. If we designed a test for this set of hypotheses whose rejection region intersected with Ω_0, reasonable persons would have cause to criticize it. Viewed from a Bayesian perspective, an observation in Ω_0 should strengthen our belief that the null hypothesis is true. Therefore observations "close to" Ω_0 should be strong evidence in support of the null hypothesis.

Thus requiring that the acceptance region contain Ω_0 means that if $\vec{p} \in \Omega_0$, evaluation of $\lambda(N)$ is unnecessary; accept H_0. This creates a slight problem if one wishes to compute a P-value for an observed $\hat{\vec{p}} \in \Omega_0$: Since we do not evaluate $\lambda(N)$, we cannot compute its P-value. However this P-value is bounded below by the quantity

$$P_{\vec{p}_0}\left\{\hat{\vec{p}} \in \Omega_0^c\right\} = P_{\vec{p}_0}\left\{\lambda(N) \leq 1\right\},$$

which can be computed without the evaluation of $\lambda(N)$.

The GLRT versus Union-Intersection Tests

In this testing situation, a union-intersection test has the decided advantage of ease of computation. However there are at times many candidate union- intersection tests, making the selection of the " best " one difficult. In fact it can be shown that the rejection region, $\mathcal{R}_\lambda$, of the GLRT is approximated by such tests in the sense that there exist at least two such union-intersection tests, whose rejection regions $\mathcal{R}_1$ and $\mathcal{R}_2$ satisfy $\mathcal{R}_1 \supseteq \mathcal{R}_\lambda \supseteq \mathcal{R}_2$. This approximation can be made " optimally"; it can be shown that there is a union-intersection test whose power is less than or equal to the power of the GLRT but greater than or equal to the power of any other union- intersection test whose rejection region is contained in $\mathcal{R}_\lambda$. This test would be a reasonable substitute for the GLRT since it should have approximately equal power and size and be much easier to compute than the GLRT statistic $\lambda(N)$. However, this "approximating" union-intersection test may not be the "natural" choice among several competing union- intersection tests, especially when $K \geq 3$. Without the aid of computing the rejection region of the GLRT described here, there is no obvious criteria for selecting it from among its competitors.

As illustrated by the following two examples, the decision whether to go to the added trouble and expense of the GLRT depends on the set of quantiles being tested.

Figures 3 and 4 compare the GLRT versus a "natural" union-intersection test for testing the hypothesis

$$H_0: \; p_1 \geq .25, \;\; p_1 + p_2 \geq .75, \;\; \text{and } p_1 + p_2 + p_3 \geq .95, \text{ with } N = 10.$$

The GLRT rejects H_0 if $\log(\lambda(N)) \geq 5.9$ with a size $\alpha = .052$.

In order to construct the Union-Intersection Test, we reproduce a portion of the Binomial Tables for N=10. The entries are the cumulative probabilities of the binomial random variable with parameter p.

Table 2: Abbreviated Binomial Table for $N = 10$.

$n_0 \setminus p$	.25	.75	.95
0	.0563	.0000	.0000
1	.2440	.0000	.0000
2	.5256	.0004	.0000
3	.7759	.0035	.0000
4	.9219	.0197	.0000
5	.9803	.0781	.0001
6	.9965	.2241	.0010
7	.9996	.4744	.0115
8	1.0000	.7560	.0861

Denote any union-intersection test that tests H_0 by

$$(.25,\ .75,\ .95,\ n_1^0,\ n_2^0,\ n_3^0;\ N),$$

indicating that the rejection region is

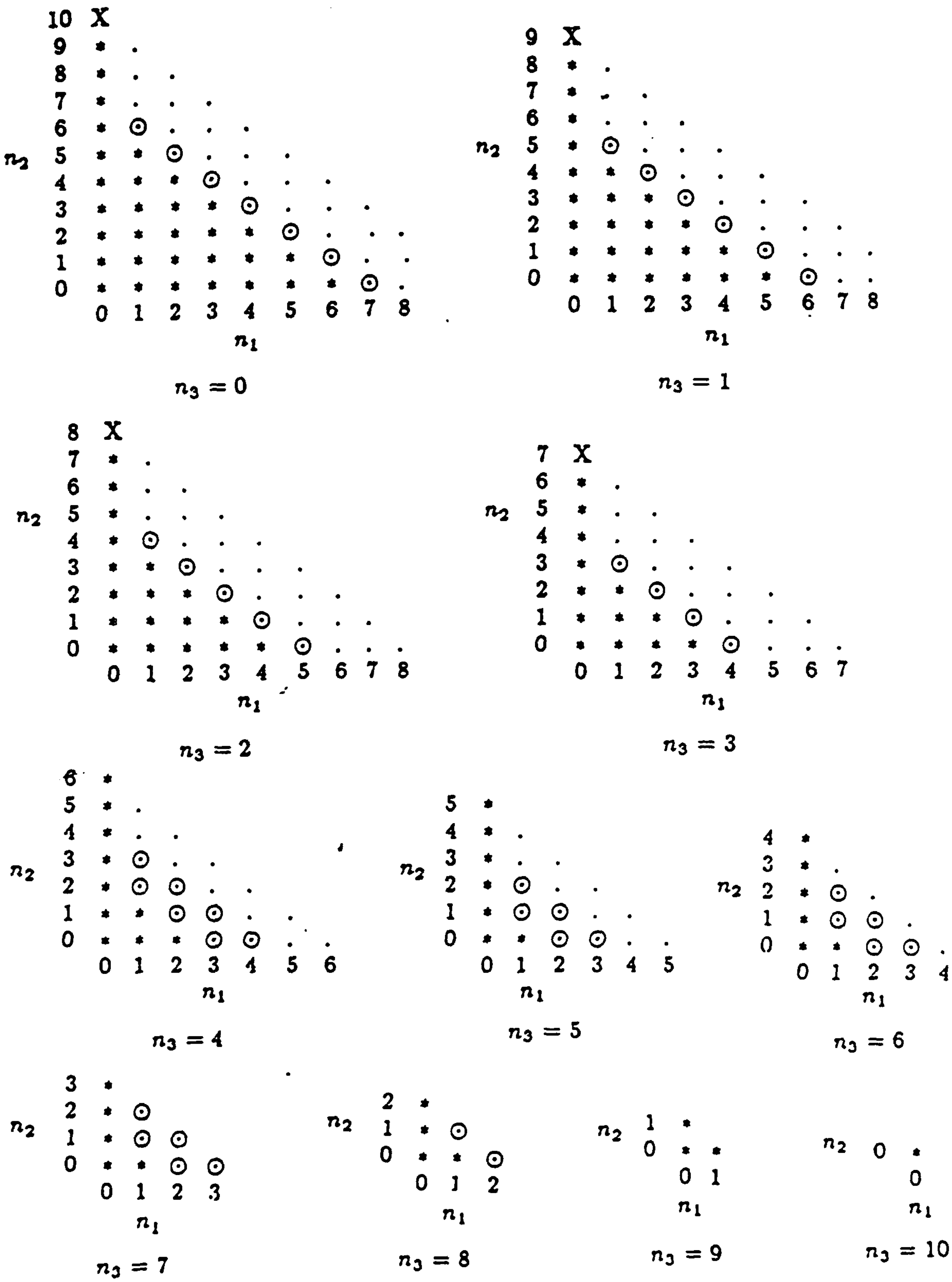
$$\{n_1 \leq n_1^0\} \cup \{n_1 + n_2 \leq n_2^0\} \cup \{n_1 + n_2 + n_3 \leq n_3^0\}.$$

We select (.25, .75, .95, 0, 1, 6; 10), the union-intersection test that rejects if $n_1 = 0$, $n_1 + n_2 \leq 1$, or $n_1 + n_2 + n_3 \leq 6$. The Bonferroni upper bound on its size is .0573. Its true size is .05718.

Note some other competitors: n_3^0 might be chosen to be any number between 2 and 7.

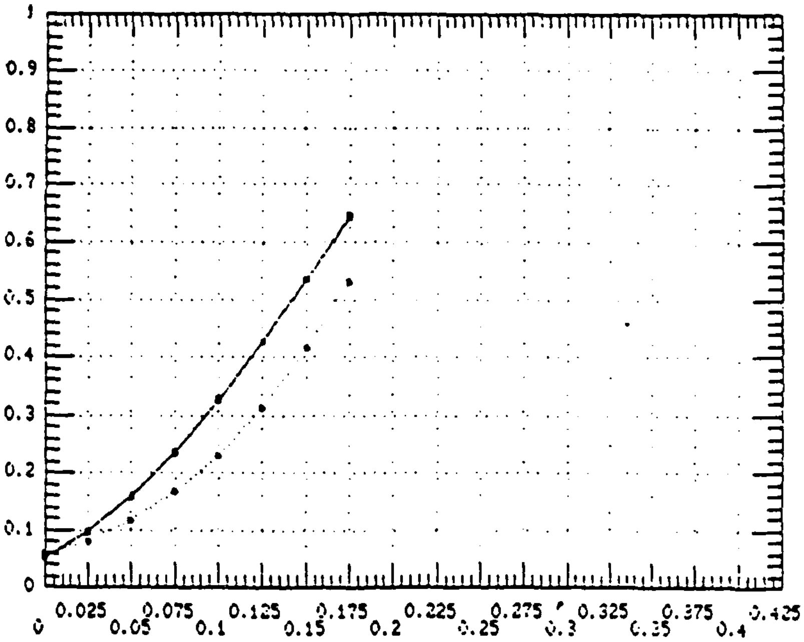
Figure 3 Rejection Regions of the GLRT and the Union-Intersection Test
(.25, .75, .95, 0, 1, 6; 10)

Note: The eleven graphs that comprise Figure 3 compare the rejection regions of the two tests. The rejection regions are given for each level of n_3. The symbols on the graphs have the following significance:
Dots ($\cdot$) mark the points not in the rejection region of either test.
Asterisks ($*$) mark points that are common to both rejection regions.
O-dots ($\odot$) mark points that are in the rejection region of the GLRT but not in the rejection region of the union-intersection test.
X's mark points that are in the rejection region of the union-intersection test but not in the rejection region of the GLRT.

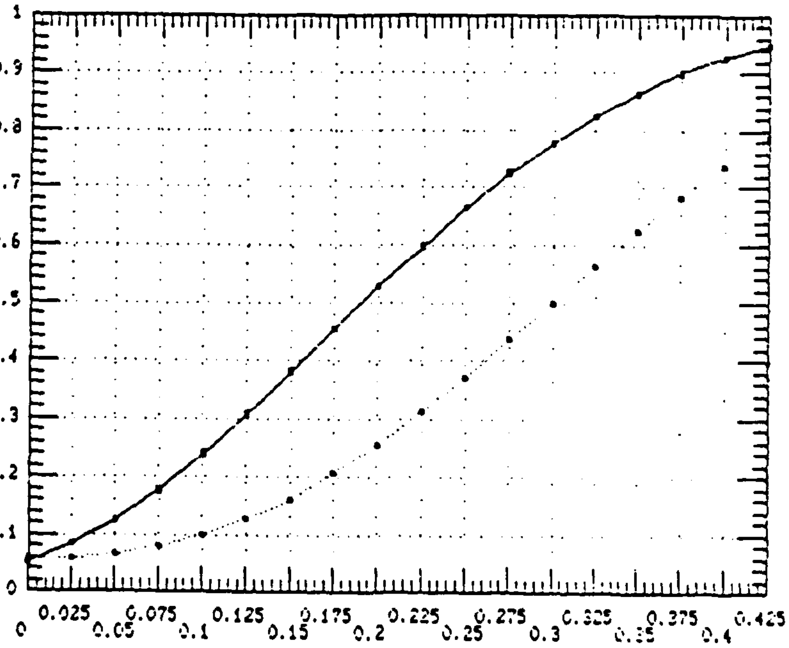

n_2
n_1
$n_3 = 0$
$n_3 = 1$
$n_3 = 2$
$n_3 = 3$
$n_3 = 4$
$n_3 = 5$
$n_3 = 6$
$n_3 = 7$
$n_3 = 8$
$n_3 = 9$
$n_3 = 10$

Figures 4 (a) - (g) Powers of the GLRT and the Union-Intersection Test (.25, .75, .95, 0, 1, 6; 10) Compared

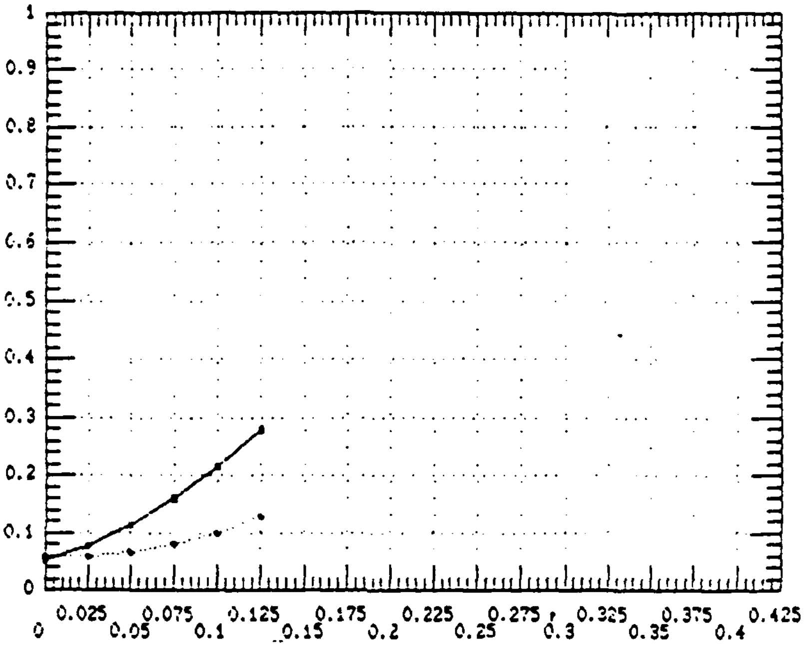
Note: Comparisons of the powers of the two tests are made at selected points along rays leaving the point $\vec{p}_0 = (.25, .5, .2)$ toward Ω_0^c. Each ray takes the form $\vec{p} = \vec{p}_0 - \gamma \vec{d}$. Here $\gamma \geq 0$ is the abscissa. The power is graphed as the ordinate. The direction vector, $\vec{d}$, is an element of the subset of R^3 that consists of the $2^3 - 1 = 7$ nontrivial vectors of the form (d_1, d_2, d_3) where $d_i = 0$ or 1, $i = 1, 2$, and 3. Points in Ω_0 were taken along these seven rays for values of $\gamma = .025$, .05, .075, .1, etc. until the smallest of the three coordinates of a selected point was less than or equal to .05. Points with at least one coordinate as small as .05 are near the boundary of Ω and into areas of extreme alternatives of little interest.
The solid line is the power of the GLRT.
The dotted line is the power of the union-intersection test.



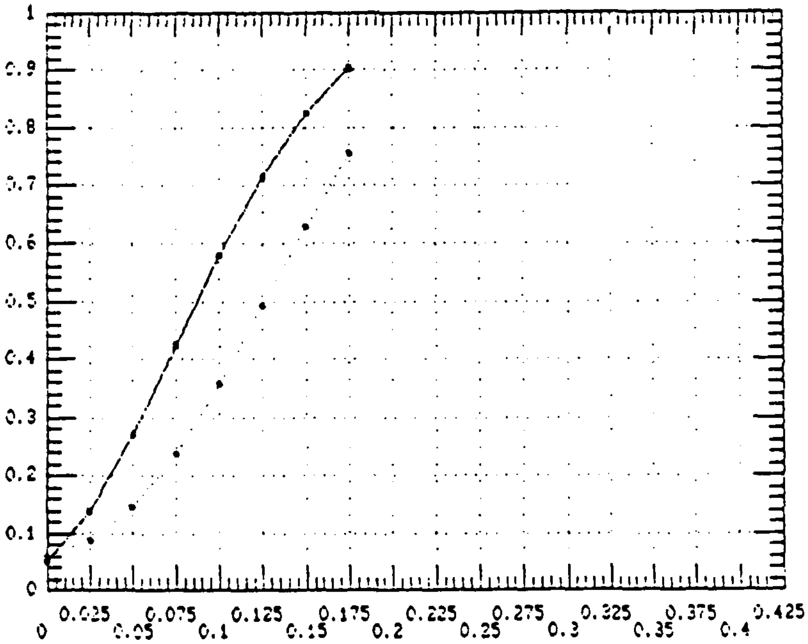
(a) $\vec{d} = (1,0,0)$. Points are of the form $(.25 - \gamma,\ .5,\ .2)$.



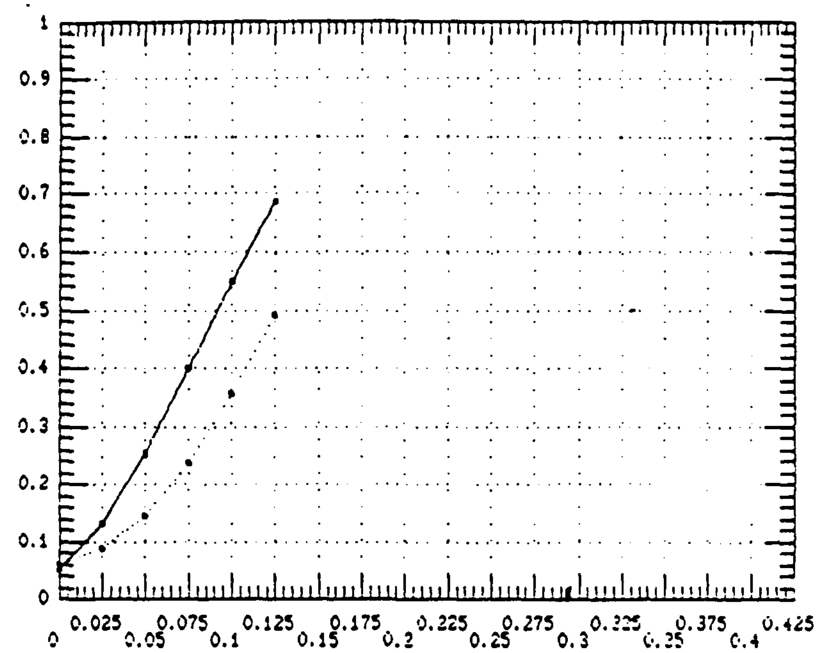
(b) $\vec{d} = (0,1,0)$. Points are of the form $(.25,\ .5 - \gamma,\ .2)$.



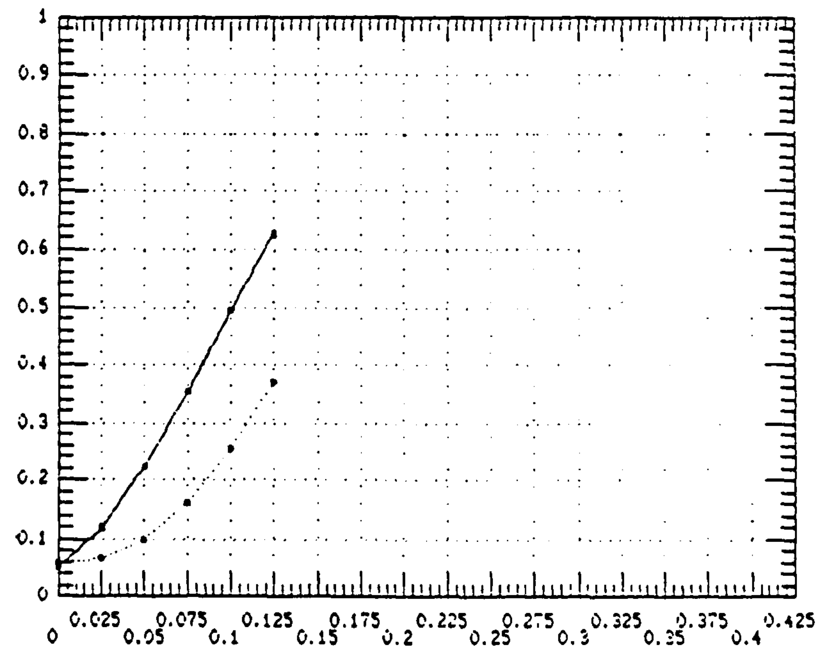
(c) $\vec{d} = (0,0,1)$. Points are of the form $(.25,\ .5,\ .2-\gamma)$.



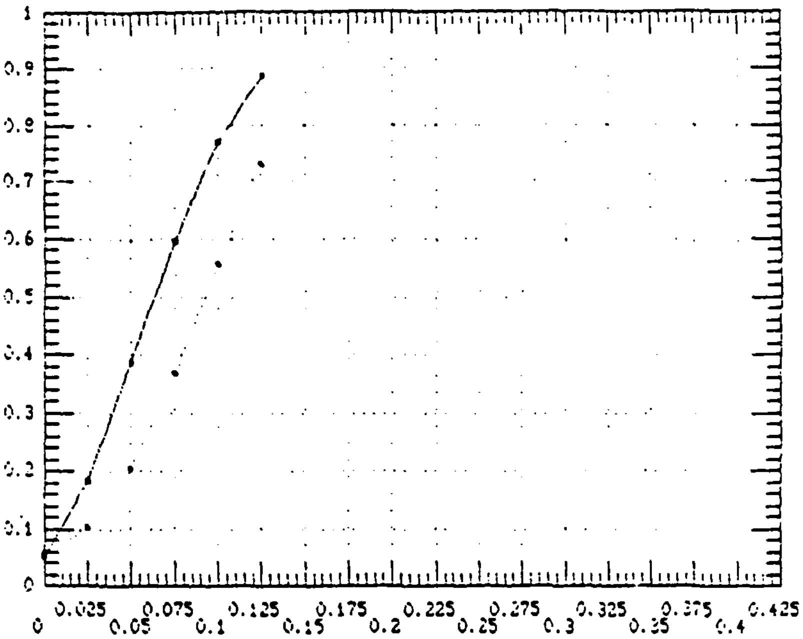
(d) $\vec{d} = (1,1,0)$. Points are of the form $(.25-\gamma,\ .5-\gamma,\ .2)$.



(e) $\vec{d} = (1,0,1)$. Points are of the form $(.25 - \gamma,\ .5,\ .2 - \gamma)$.



(f) $\vec{d} = (0,1,1)$. Points are of the form $(.25,\ .5 - \gamma,\ .2 - \gamma)$.



(g) $\vec{d} = (1,1,1)$. Points are of the form $(.25-\gamma,\ .5-\gamma,\ .2-\gamma)$.

We see from Figures 4 (a) - (g) that the power of the GLRT is larger at every point considered and substantially larger at most points.

The test (.25, .75, .95, 0, 1, 6; 10) is not the "optimal" union- intersection test whose power is approximately that of the GLRT. That test rejects if $n_1+n_2 \leq 3$, or $n_1+n_2+n_3 \leq 7$. It does not consider values of n_1 as part of its rejection region. From Table 2, we see that the upper bound on its size is .0150. An investigator would not choose this test if he or she were constructing a union-intersection test with .05 as an approximate size.

The selection of the particular set of $\vec{d}'s$ to be used in Figure 4 was not arbitrary. The vectors used in (a),(b), and (c) can be interpreted as alternatives that describe the 'breakdown' of the null hypothesis in only one quantile. For example, (a) are points in Ω_0^c that fail to be in Ω_0 because of their first coordinates; the "interquantile range" is adequate for the second and third quantiles. These distributions fall short only on the first specified quantile. Likewise (c), (d), and (e) are points that fail to be members of Ω_0 because of two coordinates. This is a more pathological situation than the situation in (a), (b), or (c). Figure 4 (g), in the sense of ordering described so far, represents the most extreme case of the failure of the null hypothesis to be true. In this case, the distribution comes up short on all three quantiles. The graphs indicate that, for the same γ, both the GLRT and the union- intersection test have higher power on (g) than on (d), (e), or (f) and higher power on these three than on (a), (b), and (c). In this sense the power of both tests is sensitive to the number of parameters that fail to meet the requirements of the null hypothesis.

All of the graphs in Figure 4 except (b) appear incomplete. This is because rays from $\vec{p_0}$ can extend for longer distances and remain in Ω for some $\vec{d}$ than for others.

For testing

$$\mathrm{H}_0: \; p_1 \geq .25, \; p_1+p_2 \geq .75, \quad \text{and } p_1+p_2+p_3 \geq .95,$$

Figures 4 (a)-(g) illustrated an advantage of the GLRT over the union-intersection test (.25, .75, .95, 0, 1, 6; 10) with respect to power. Not only is the power larger in every case, but the Type I error is closer to the specified value.

The example that follows illustrates a case where the GLRT serves only to "fine tune" the Type I error rate of the union-intersection test rather than to improve its power appreciably. Consider testing

$$\mathrm{H}_0: \; p_1 \geq .3, \; p_1+p_2 \geq .6, \quad \text{and } p_1+p_2+p_3 \geq .9.$$

The GLRT rejects if $-2\log(\lambda(N)) \leq 4.8$ with a size $\alpha = .0503$. We display Table 3 in order to illustrate the selection process for

$$(.3, .6, .9, n_1^0, n_2^0, n_3^0; 10).$$

Table 3: Abbreviated Binomial Table for $N = 10$.

$n_0 \setminus p$	.30	.60	.90
0	.0282	.0001	.0000
1	.1493	.0017	.0000
2	.3828	.0123	.0000
3	.6496	.0548	.0000
4	.8497	.1662	.0001
5	.9527	.3669	.0016
6	.9894	.6177	.0128
7	.9984	.8327	.0702
8	.9999	.9536	.2639

We select the union-intersection test (.3, .6, .9, 0, 2, 6; 10). The upper bound on its size is .053. Its true size is .048. This is the union-intersection test described earlier whose rejection region is contained in the rejection region of the GLRT and has the highest power of any union- intersection test with $\mathcal{R} \subseteq \mathcal{R}_\lambda$.

The rejection regions of these two tests are identical except that the GLRT rejects H_0 if $(n_1, n_2, n_3) = (1, 2, 4)$, while the union-intersection rejection test does not. The powers of these two tests will obviously be similar except for alternatives close to this point and even for such alternatives, the powers only differ by approximately .00025, the mass of that point when $(p_1, p_2, p_3) = (.1, .2, .4)$. The marginal improvement that is the result of using the GLRT is to increase the Type I error rate to almost exactly .05.

Summary and Conclusions

The purpose of this work reported here is to construct a small sample test procedure to test the null hypothesis

$$H_0 : \sum_{\nu=1}^{i} p_\nu \geq \sum_{\nu=1}^{i} p_{0\,\nu}, \; 1 \leq i \leq K$$

which is superior to any union-intersection test based on $\sum_{\nu=1}^{i} N_\nu$, $1 \leq i \leq K$, with critical values chosen from the binomial tables. There is little computation needed to apply the union-intersection test. One simply observes n_1, $n_1 + n_2$, $\cdots$, $\sum_{\nu=1}^{K} n_\nu$. If any one of these is smaller than a corresponding predetermined critical value, the null hypothesis is rejected. The Type I error rate of this test can only be estimated from above by the binomial tables. For $K \geq 3$ there are usually several different combinations of critical values from which to choose. The tables give no criteria upon which to base a decision as to which combination is the best choice. Of course one can compute the multinomial probabilities upon which the Type I error rate of a test based on a particular set of critical values depends, but there is no guarantee that any union-intersection test can be constructed yielding an α close to some predetermined level. Even if one is found, it may be such that $\sum_{\nu=1}^{i} n_\nu$ for some $i \leq K$ may not be considered in the decision process. This is unsettling if we keep in mind that this is supposed to be a simultaneous test on all of the quantiles.

Juxtaposed to the simplicity and uncertainty of the union-intersection test is the complexity and certainty of the GLRT. An algorithm is provided that facilitates the evaluation of the statistic. Although computation of the p-value or critical value is prohibitively complicated if done by hand, on a AT-Type PC, computations of p-values for $K = 7$ and $N = 20$ take less than 5 minutes. It has been illustrated that the relative superiority of the power of the GLRT over the union-intersection test depends on the choice of the set of critical values of the union-intersection test and the set of null hypothesis parameters. It appears that the most apparent choice of critical values for the union-intersection test can yield a test with far less power than the GLRT even though there is a union-intersection test whose power is less than but approximately the same as power of the GLRT. That test will more than likely not be chosen by the researcher because the upper bound on its Type I error rate is below the nominal rate and/or it requires excluding one or more of the quantiles from consideration. It is because of the uncertainty of the union- intersection test that, in spite the complexity of computation, the GLRT can be a useful alternative.

References

Chernoff, H. (1954) " On the distribution of the likelihood ratio", Annals of Mathematical Statistics, **25**, 573-578.

Feder, P. J. (1968) "On the distribution of the log likelihood ratio test statistic when the true parameter is 'near' the boundaries of the hypothesis regions", Annals of Mathematical Statistics, **39**, 2044-2055.

Greenberg, I. (1985) "A One-Sided Goodness-of-Fit Test for a Multinomial Population", Journal of the American Statistical Association, **80**, 558-561.

Marshal, A. and Olkin I. (1984) Theory of Majorization and its Applications, New York: Academic Press.

Roy, S. M. (1953) "On a hueristic method of test construction and its use in multivariate analysis ",Annals of Mathematical Statistics **24**, 220-238.

Self S. G. and Liang K. (1984) "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Non-Standard Conditions ", Journal of the American Statistical Association, **82**, 605-610.

Appendix A

Proofs of Results Supporting the Evaluation Algorithm

We restate Theorem 1.

<u>THEOREM 1</u>. *Define i_0 to be 0 :*

$$\hat{p}_i^{(i_1,\cdots,i_t)} = \begin{cases} \text{(a)} \dfrac{\hat{p}_i \, P_0{}_{i_{j-1}}^{i_j}}{\hat{P}_{i_{j-1}}^{i_j}}, & \text{if } i_{j-1} < i \leq i_j,\ 1 \leq j \leq t, \text{ and } \hat{P}_{i_{j-1}}^{i_j} \neq 0, \\ \text{(b)} \quad p_{0\,i}, & \text{if } i_{j-1} < i \leq i_j,\ 1 \leq j \leq t, \text{ and } \hat{P}_{i_{j-1}}^{i_j} = 0, \\ \text{(c)} \quad \dfrac{\hat{p}_i \left(1 - P_0{}^{i_t}\right)}{1 - \hat{P}^{i_t}}, & \text{if } i_t < i \leq K \text{ and } \hat{P}_0^{i_t} \neq 1, \\ \text{(d)} \quad 0, & \text{if } i_t \leq i \leq K \text{ and } \hat{P}^{i_t} = 1. \end{cases}$$

PROOF:

The case when $i_t + 1 \leq K$:

If $i_{j-1} + 1 = i_j$, write the $i_j{}^{th}$ restriction, $P^{i_j} = P_0{}^{i_j}$, as $P^{i_{j-1}} + p_{i_j} = P_0{}^{i_{j-1}} + p_{0\,i_j}$. Since restriction i_{j-1} is $P^{i_{j-1}} = P_0^{i_{j-1}}$, $\hat{P}_{i_{j-1}}^{i_j} = \hat{p}_{i_j}$ and $P_0{}_{i_{j-1}}^{i_j} = p_{0 i_j}$. Thus

$$\hat{p}_{i_j}^{(i_1,\cdots,i_t)} = p_{0\,i_j} = \frac{\hat{p}_{i_j} P_0{}_{i_{j-1}}^{i_j}}{\hat{P}_{i_{j-1}}^{i_j}}$$

and the given formula part (a) holds trivially.

Assume that for the j^{th} restriction $1 \leq j \leq t$, $i_{j-1} + 1 < i_j$. As was noted, the t restrictions placed on the coordinates of $\vec{p}$ that define $\text{Plane}_{\vec{p}_0}^{(i_1,\cdots,i_t)}$ can be written, for $j = 1, \cdots t$,

$$\sum_{\nu=i_{j-1}+1}^{i_j} p_\nu \equiv P_{i_{j-1}}^{i_j} = P_0{}_{i_{j-1}}^{i_j}.$$

Rewriting $\log L(\hat{\vec{p}}, \vec{p})$ using the t restrictions in this form we obtain

$$\sum_{j=1}^{t} \left[\sum_{\nu=i_{j-1}+1}^{i_j - 1} \hat{p}_\nu \log(p_\nu) + \hat{p}_{i_j} \log\left(P_0{}_{i_{j-1}}^{i_j} - P_{i_{j-1}}^{i_j - 1} \right) \right] + \sum_{\nu=i_t+1}^{K} \hat{p}_\nu \log(p_\nu) +$$
$$\left(1 - \hat{P}^K\right) \log\left(1 - P_0{}^{i_t} - P_{i_t}^K\right).$$

Assume also that $\hat{P}^{i_j}_{i_{j-1}} \neq 0$ as in formula part (a).
We are seeking an expression for $\hat{p}_\nu^{(i_1,\cdots,i_t)}$ for $i_{j-1}+1 < \nu \leq i_j$. The only part of $\log L(\hat{\vec{p}}, \vec{p})$ involving p_ν, for such ν, is

$$\sum_{\nu=i_{j-1}+1}^{i_j-1} \hat{p}_\nu \log(p_\nu) + \hat{p}_{i_j} \log\left({P_0}^{i_j}_{i_{j-1}} - P^{i_j-1}_{i_{j-1}}\right). \tag{A1}$$

(The parameter p_{i_j} is represented in (A1) by ${P_0}^{i_j}_{i_{j-1}} - P^{i_j-1}_{i_{j-1}}$ since the j^{th} restriction requires that ${P_0}^{i_j}_{i_{j-1}} = P^{i_j}_{i_{j-1}}$.)

Taking partial derivatives with respect to p_ν in (A1) and setting them equal to zero yields the following system of equations:

$$\frac{\hat{p}_\nu}{p_\nu} = \frac{\hat{p}_{i_j}}{{P_0}^{i_j}_{i_{j-1}} - P^{i_j-1}_{i_{j-1}}}, \quad \nu = i_{j-1}+1, \cdots, i_j - 1.$$

It follows that

$$\hat{p}_\nu^{(i_1,\cdots,i_t)} = \frac{\hat{p}_\nu {P_0}^{i_j}_{i_{j-1}}}{\hat{P}^{i_j}_{i_{j-1}}}, \quad \nu = i_{j-1}+1, \cdots, i_j.$$

In the special case that $\hat{P}^{i_j}_{i_{j-1}} = 0$, as in formula part (b), expression (A1) is identically 0, i.e., the restricted likelihood function is not a function of p_ν for $i_{j-1}+1 \leq \nu \leq i_j$ whose values have no effect in maximizing the likelihood function on $\text{Plane}_{\vec{p}_0}^{(i_1,\cdots,i_t)}$. The restriction on $\hat{p}_\nu^{(i_1,\cdots,i_t)}$ for $i_{j-1}+1 \leq \nu \leq i_j$ is that $\sum_{i_{j-1}+1}^{i_j} \hat{p}_\nu^{(i_1,\cdots,i_t)} = {P_0}^{i_j}_{i_{j-1}}$. Thus, for consistency we may define $\hat{p}_\nu^{(i_1,\cdots,i_t)} = p_{0\,\nu}$ for such ν.

Now let $i_t + 1 \leq \nu$ and assume $\hat{P}^{i_t} \neq 1$ as in formula part (c).
The portion of $\log L(\hat{\vec{p}}, \vec{p})$ involving such p_ν is

$$\sum_{\nu=i_t+1}^{K} \hat{p}_\nu \log(p_\nu) + \left(1 - \hat{P}^K\right) \log\left(1 - {P_0}^{i_t} - P^K_{i_t}\right).$$

Again taking partial derivatives with respect to p_ν for $i_t + 1 \leq \nu \leq K$ and setting them equal to zero yields the system of equations

$$\frac{\hat{p}_\nu}{p_\nu} = \frac{1 - \hat{P}^K}{1 - {P_0}^{i_t} - P^K_{i_t}}.$$

It follows that

$$\hat{p}_\nu^{(i_1,\cdots,i_t)} = \frac{\hat{p}_\nu\left(1 - P_0^{i_t}\right)}{1 - \hat{P}^{i_t}}.$$

In the case where $\hat{P}^{i_t} = 1$, as in formula part (d), $\hat{p}_\nu = 0$ for $i_t + 1 \leq \nu \leq K$. Since there are no restrictions on these parameters, we use the global MLE's and define $\hat{p}_\nu^{(i_1,\cdots,i_t)} = \hat{p}_\nu = 0$ for such ν.

The case when $i_t = K$:

In this case the expression for $\log L(\hat{\vec{p}}, \vec{p})$, employing the t restrictions, simplifies to

$$\sum_{j=1}^{t}\left[\sum_{\nu=i_{j-1}+1}^{i_j-1} \hat{p}_\nu \log(p_\nu) + \hat{p}_{i_j} \log\left(P_0{}_{i_{j-1}}^{i_j} - P_{i_{j-1}}^{i_j-1}\right)\right] +$$
$$\left(1 - \hat{P}^K\right)\log\left(1 - P_0{}^K\right).$$

The case $i_t + 1 \leq \nu$, the conditions for tha application of formula parts (c) and (d), do not occur. Thus the previous argument for $i_{t-1} + 1 \leq \nu \leq i_t = K$ holds here. $\diamond$

The following theorem validates the Evaluation Algorithm:

<u>THEOREM 2</u>. *If $\hat{p} \in \Omega_0^c$ then $\hat{p}_0 = \hat{\vec{p}}^{(i_1,\cdots,i_t)}$, where $\{i_1,\cdots,i_t\}$ is the smallest subset of $\{1,2,\cdots,K\}$ such that $\hat{\vec{p}}^{(i_1,\cdots,i_t)} \in \Omega_0$.*

This follows from the following sequence of lemmas.

<u>LEMMA 1A</u>. *If $\hat{\vec{p}}^{(i_1)} \in \Omega_0$ and $\hat{\vec{p}} \notin \Omega_0$, then*

$$\hat{P}^{i_1} \leq P_0{}^{i_1}.$$

PROOF: We are given that $\hat{\vec{p}}^{(i_1)} \in \Omega_0$ or, $\hat{P}^{(i_1)l} \geq P_0{}^l$, $\forall l$. But the assumption that $\hat{p} \notin \Omega_0$ implies that for at least one l, $\hat{P}^l < P_0{}^l$:

(1) If $l = i_1$ and $\hat{P}^{i_1} < P_0{}^{i_1}$, we have the result.

(2) If $l \leq i_1 - 1$ and $\hat{P}^l < P_0{}^l$, for such l from Theorem 1 formula parts (a) and (b)

$$\hat{P}^{(i_1)l} \geq P_0{}^l \iff \hat{P}^l P_0{}^{i_1} \geq P_0{}^l \hat{P}^{i_1}.$$

The assumption that $\hat{P}^l < P_0{}^l$ implies that

$$\hat{P}^l P_0{}^{i_1} > \hat{P}^l \hat{P}^{i_1}.$$

The assumptions that $\hat{P}^{(i_1)l} \geq P_0{}^l$ and that $\hat{P}^{i_1} \neq 0$ assure that $\hat{P}^l \neq 0$. This permits us to conclude that

$$\hat{P}^{i_1} < P_0{}^{i_1}.$$

(3) If $i_1 < l$, then from Theorem 1 formula parts (c) and (d),

$$\hat{P}^{(i_1)l} \geq P_0{}^l \iff \hat{P}^l_{i_1}\left(1 - P_0{}^{i_1}\right) \geq P_0{}^l_{i_1}\left(1 - \hat{P}^{i_1}\right).$$

Subtract the term $P^l_{0i_1}\hat{P}^l_{i_1}$ from both sides of the above expression to obtain

$$\hat{P}^{(i_1)l} \geq P_0{}^l \iff \hat{P}^l_{i_1}\left(1 - P_0{}^l\right) \geq P_0{}^l_{i_1}\left(1 - \hat{P}^l\right).$$

Recall that $\hat{P}^l < P_0{}^l$ or $1 - P_0{}^l < 1 - \hat{P}^l$. Thus

$$\hat{P}^l_{i_1}\left(1 - P_0{}^l\right) \geq P_0{}^l_{i_1}\left(1 - \hat{P}^l\right) > P_0{}^l_{i_1}\left(1 - P_0{}^l\right).$$

Considering only the extreme terms and dividing both sides by $1 - P_0{}^l$ we obtain $\hat{P}^l_{i_1} > P_0{}^l_{i_1}$. But if $\hat{P}^l < P_0{}^l$, then $\hat{P}^{i_1} < P_0{}^{i_1}$. ◇

If we denote $\vec{p}$ as $\vec{p}^{(i_0)}$, the superscript i_0 denoting that it is the '0 restriction MLE', then we may restate Lemma 1A as:

If $\hat{\vec{p}}^{(i_0,i_1)} \in \Omega_0$ and $\hat{\vec{p}}^{(i_0)} \notin \Omega_0$, then

$$\hat{P}^{(i_1)i_1} \leq P_0{}^{i_1}.$$

In this context, the following lemma is the analog of Lemma 1A when we treat a 't restriction MLE', $\hat{\vec{p}}^{(i_1,\cdots,i_t)}$.

<u>LEMMA 2A</u>. *Let $t \geq 2$ and assume that $\hat{\vec{p}}^{(i_1,\cdots,i_t)} \in \Omega_0$.*

(1) If $\hat{\vec{p}}^{(i_1,\cdots,i_{t-1})} \notin \Omega_0$, then $\hat{P}^{(i_1 i_2 \cdots i_{t-1})i_t} < P_0{}^{i_t}$.

(2) If $\hat{\vec{p}}^{(i_2,i_3\cdots,i_t)} \notin \Omega_0$, then $\hat{P}^{(i_2\cdots i_t)i_1}_{1} < P_0{}^{i_1}$

(3) If for any j, $2 \leq j \leq t-1$, $\hat{\vec{p}}^{(i_1,\cdots,i_{j-1},i_{j+1},\cdots,i_t)} \notin \Omega_0$, then

$$\hat{P}^{(i_1\cdots i_{j-1},i_{j+1},\cdots,i_t)i_j} < P_0{}^{i_j}$$

The proof of this lemma is similar to the proof of Lemma 1A and will be omitted.

<u>LEMMA 3A</u>. *If the number of restrictions $t \geq 1$ and $\hat{\vec{p}}^{(i_1,\cdots,i_t)} \in \Omega_0$ but none of the edges characterized by any of the t subsets of order t-1 of the set $\{i_1,\cdots,i_t\}$ contain their restricted MLE's, then*

$$\hat{P}^{i_j}_{i_{j-1}}\left(1 - P_0{}^{i_t}\right) < P_0{}^{i_j}_{i_{j-1}}\left(1 - \hat{P}^{i_t}\right), \quad 1 \leq j \leq t. \qquad (A2)$$

PROOF: Note that the assumptions of this lemma are the same as the assumptions of Lemmas 1A and 2A. We can apply the results of those lemmas to state that $\hat{P}^{(i_2\cdots i_t)i_1} < P_0{}^{i_1}$,

$\hat{P}^{(i_1 i_3 \cdots i_t) i_2} < P_0{}^{i_2}, \ldots \hat{P}^{(i_1 i_2 \cdots i_{t-2} i_t) i_{t-1}} < P_0{}^{i_{t-1}}$, and $\hat{P}^{(i_1 i_2 \cdots i_{t-1}) i_t} < P_0{}^{i_t}$ all must be true.

(1) Start an induction argument with $j = t$.

Because $\hat{P}^{(i_1 i_2 \cdots i_{t-1}) i_{t-1}} = P_0{}^{i_{t-1}}$,

$$\hat{P}^{(i_1 i_2 \cdots i_{t-1}) i_t} < P_0{}^{i_t} \iff \hat{P}^{(i_1 i_2 \cdots i_{t-1}) i_t}_{i_{t-1}} < P_0{}^{i_t}_{i_{t-1}}$$

$$\iff \hat{P}^{i_t}_{i_{t-1}} \left(1 - P_0{}^{i_{t-1}}\right) < P_0{}^{i_t}_{i_{t-1}} \left(1 - \hat{P}^{i_{t-1}}\right),$$

from Theorem 1 formula part (c). Subtracting $\hat{P}^{i_t}_{i_{t-1}} P_0{}^{i_t}_{i_{t-1}}$ from both sides of the last expression, we have

$$\hat{P}^{i_t}_{i_{t-1}} \left(1 - P_0{}^{i_t}\right) < P_0{}^{i_t}_{i_{t-1}} \left(1 - \hat{P}^{i_t}\right).$$

This theorem holds for $j = t$.

(2) Now we argue by "reverse" induction; we assume that the result holds for $j + 1$ and show that it holds for j.

If it is true for $j + 1$ we have

$$\hat{P}^{i_{j+1}}_{i_j} \left(1 - P_0{}^{i_t}\right) < P_0{}^{i_{j+1}}_{i_j} \left(1 - \hat{P}^{i_t}\right). \tag{A3}$$

Recall that

$$\hat{P}^{(i_1 \cdots i_{j-1} i_{j+1} \cdots i_t) i_j} \leq P_0{}^{i_j} \iff \hat{P}^{(i_{j-1} i_{j+1}) i_j} \leq P_0{}^{i_j}_{i_{j-1}} \iff$$

$$\hat{P}^{i_j}_{i_{j-1}} P_0{}^{i_{j+1}}_{i_{j-1}} < P_0{}^{i_j}_{i_{j-1}} \hat{P}^{i_{j+1}}_{i_{j-1}}.$$

Subtracting $\hat{P}^{i_j}_{i_{j-1}} P_0{}^{i_j}_{i_{j-1}}$ from both sides we write

$$\hat{P}^{(i_{j-1} i_j) i_j} \leq P_0{}^{i_j}_{i_{j-1}} \iff \hat{P}^{i_j}_{i_{j-1}} P_0{}^{i_{j+1}}_{i_j} < P_0{}^{i_j}_{i_{j-1}} \hat{P}^{i_{j+1}}_{i_j}.$$

Now multiply (A3) by $P_0{}^{i_j}_{i_{j-1}}$ and apply this last expression of $\hat{P}^{(i_{j-1} i_j) i_j}$ to obtain

$$\hat{P}^{i_j}_{i_{j-1}} P_0{}^{i_{j+1}}_{i_j} \left(1 - P_0{}^{i_t}\right) < P_0{}^{i_j}_{i_{j-1}} \hat{P}^{i_{j+1}}_{i_j} \left(1 - P_0{}^{i_t}\right) < P_0{}^{i_j}_{i_{j-1}} P_0{}^{i_{j+1}}_{i_j} \left(1 - \hat{P}^{i_t}\right).$$

The extremes of the above inequality with both sides divided by $P_0{}^{i_{j+1}}_{i_j}$ yield

$$\hat{P}^{i_j}_{i_{j-1}} \left(1 - P_0{}^{i_t}\right) < P_0{}^{i_j}_{i_{j-1}} \left(1 - \hat{P}^{i_t}\right).$$

This is the desired result for j. By induction the result holds for $1 \le j \le t$. $\diamond$

We now have established the prerequisites to prove Theorem 2, the theorem that supports the evaluation algorithm.

PROOF OF THEOREM 2: (1) Assume first that $t = 1$, $\hat{p} \in \Omega_0^c$, and $\hat{\vec{p}}^{(i_1)} \in \Omega_0$. We will prove that

$$\max_{\vec{p}\in\Omega_0} \log L(\hat{\vec{p}}, \vec{p}) = \log L(\hat{\vec{p}}, \hat{\vec{p}}^{(i_1)}).$$

Written with respect to terms involve ${P_0}^{i_1}$

$$\log L(\hat{\vec{p}}, \hat{\vec{p}}^{(i_1)}) \equiv \hat{P}^{i_1} \log({P_0}^{i_1}) + (1 - \hat{P}^{i_1}) \log(1 - {P_0}^{i_1}) + C = h({P_0}^{i_1}) + C.$$

Here C represents all those terms in $\log L(\hat{\vec{p}}, \hat{\vec{p}}^{(i_1)})$ that do not depend on ${P_0}^{i_1}$. Taking the derivative of h(.) with respect to ${P_0}^{i_1}$ leads to the conclusion that the maximum value of the likelihood function restricted to the plane satisfying the restriction $P^{i_1} = \rho$ decreases with increasing values of ρ on the interval $\rho \ge \hat{P}^{i_1}$. Now consider any point $\vec{p}^* \in \Omega_0$. By the definition of Ω_0, the first i_1 coordinates of $\vec{p}^*$ must satisfy ${P_0}^{i_1} \le \sum_{\nu=1}^{i_1} p_\nu^* = P^{*i_1}$. Thus $\vec{p}^*$ lies on a plane parallel to $\text{Plane}_{\vec{p}_0}^{(i_1)}$ but with elements satisfying the restriction $P^{i_1} = P^{*i_1}$. Because of this the maximum value of the likelihood function restricted to the plane satisfying the restriction $P^{i_1} = P^{*i_1}$ is not larger than $\log L(\hat{\vec{p}}, \hat{\vec{p}}^{(i_1)})$. But $\vec{p}^*$ is an element of that parallel plane and thus $\log L(\hat{\vec{p}}, \vec{p}^*) \le \log L(\hat{\vec{p}}, \hat{\vec{p}}^{(i_1)})$. We conclude that $\hat{\vec{p}}^{(i_1)}$ maximizes $\log L(\hat{\vec{p}}, \vec{p})$ as a function of $\vec{p}$ on Ω_0. The theorem is proved for this case if $\hat{P}^{i_1} \le {P_0}^{i_1}$ but this follows from Lemma 1A.

Now consider the case when $\hat{\vec{p}}^{(i_1,\cdots,i_t)} \in \Omega_0$, $t \ge 2$, and none of the MLE's restricted to "planes" characterized by the t subsets of $\{i_1, \cdots, i_t\}$ with $t-1$ elements are in Ω_0. We will show that the log-likelihood function, $\log L(\hat{\vec{p}}, \vec{p})$, is maximized on Ω_0 by $\hat{\vec{p}}^{(i_1,\cdots,i_t)}$. Proceeding in a similar fashion as the case above, we write

$$\log L(\hat{\vec{p}}, \hat{p}^{(i_1,\cdots,i_t)}) =$$

$$\sum_{j=1}^{t} \hat{P}_{i_{j-1}}^{i_j} \log({P_0}_{i_{j-1}}^{i_j}) + (1 - \hat{P}^{i_t}) \log(1 - {P_0}^{i_t}) + \text{ C }.$$

The last term, C, is a constant with respect to ${P_0}_{i_{j-1}}^{i_j}$ for $1 \le j \le t$. For any $\vec{p} \in \Omega$, denote $\rho_j = P_{i_{j-1}}^{i_j}$ for $1 \le j \le t$. Consider the above expression as a function of the ρ_j;

$$f(\rho_1, \rho_2, \cdots, \rho_t) = \sum_{j=1}^{t} (\hat{P}_{i_{j-1}}^{i_j}) \log(\rho_j) + (1 - \hat{P}^{i_t}) \log\left(1 - \sum_{j=1}^{t} \rho_j\right).$$

Letting $\rho_j^0 = P_{0\,i_{j-1}}^{\;i_j}$,

$$\log L(\hat{\vec{p}}, \hat{p}_\nu^{(i_1,\cdots,i_t)}) = f(\rho_1^0, \rho_2^0, \cdots, \rho_t^0) + \ C\ .$$

The partial derivatives of $f(\cdot)$ with respect to ρ_j, $1 \le j \le t$, yield the following conditions:

$$\frac{\partial f}{\partial \rho_j} < 0 \iff \frac{\hat{P}_{i_{j-1}}^{i_j}}{\rho_j} < \frac{1 - \hat{P}^{i_t}}{1 - \sum_{\nu=1}^{t} \rho_\nu} \iff$$

$$\hat{P}_{i_{j-1}}^{i_j} \left(1 - \sum_{\nu=1}^{t} \rho_\nu\right) < \rho_j \left(1 - \hat{P}^{i_t}\right) \ \text{ for } \ 1 \le j \le t. \tag{A4}$$

Note that if $\rho_j = \rho_j^0$, then (A4) is exactly (A2) and (A2) is true under the given conditions. Define

$$\mathcal{D}_{\vec{\rho}_0} = \{\vec{\rho} \in R^t \mid \rho_j \ge \rho_j^0,\ \forall j,\ 1 \le j \le t\}.$$

Inequality (A4) is true for any $\vec{\rho} \in \mathcal{D}_{\vec{\rho}_0}$. Choose any $\vec{p}^* \in \Omega_0$ and let

$$\rho_j^* = \sum_{\nu = i_{j-1}+1}^{i_j} p_\nu^*, \quad 1 \le j \le t.$$

Note that $\vec{\rho}^* \in \mathcal{D}_{\vec{\rho}_0}$. From (A4)

$$\log L(\hat{\vec{p}}, \hat{p}_\nu^{(i_1,\cdots,i_t)}) = f(\rho_1^0, \rho_2^0, \cdots, \rho_t^0) + \ C \ \ge$$

$$f(\rho_1^*, \rho_2^0, \cdots, \rho_t^0) + \ C \ \ge f(\rho_1^*, \rho_2^*, \cdots, \rho_t^0) + \ C \ \ge \cdots \ge$$

$$f(\rho_1^*, \rho_2^*, \cdots, \rho_t^*) + \ C \ = \log L(\hat{\vec{p}}, \vec{p}^*).$$

Since $\vec{p}^*$ was an arbitrary element in Ω_0, we have proven the theorem. $\diamond$

Appendix B

Proofs of Results Supporting the Approximation of the Rejection Region of the GLRT with Rejection Regions of Union-Intersection Tests

Let $\vec{p} \in \mathcal{R}^K$, $0 < p_i$, and $\sum_{\nu=1}^{K} p_i \leq 1$.

DEFINITION.

(1) Denote by $T(p_i)$ the univariate binomial test with acceptance region

$$\left\{\hat{\vec{p}} \,\middle|\, \hat{P}^i \geq p_i\right\}.$$

(2) The union-intersection test whose acceptance region is

$$\bigcap_{i=1}^{K}\left\{\hat{\vec{p}} \,\middle|\, \hat{P}^i \geq p_i\right\} \quad \text{is then denoted} \quad \bigcap_{i=1}^{K} T(p_i).$$

It is well known that for each i, $T(p_i)$ is the uniformly most powerful unbiased test of its size for testing the hypothesis

$$\mathrm{H}_0^i : P^i \geq P_0{}^i.$$

Arguments that follow will show that the rejection region of the GLRT can always be bounded above and below in the sense of set inclusion by the rejection regions of two union-intersection tests of the type defined in the definition (2) above. This result has consequences both for the determination of the acceptance region of the likelihood ratio test discussed here and for power comparisons with other tests.

Lemma 1B is used to prove Lemma 2B. Each are proven are proven similarly to Lemma 2A of Appendix A and thus are only stated here. Lemma 2B and 3B are used to prove Lemma 4B, the central result used in the identification of the lower-bounding Union-Intersection Test.

LEMMA 1B. *Under the conditions of Lemma 1A or 2A (Appendix A), we have*

$$\hat{p}^{(i_{j-1})i_j}, \leq P_0{}^{i_j} \quad j = 2, \cdots, t.$$

LEMMA 2B.

Under the conditions of Lemma 1A or 2A (Appendix A), $\hat{P}^{i_j} < P_0{}^{i_j}$, for $1 \leq j \leq t$.

This next lemma describes the behavior of the univariate likelihood functions of the $\sum_{\nu=1}^{i} n_\nu$, $1 \leq i \leq K$.

LEMMA 3B. *The function*

$$g_a(x) = x\log\left(\frac{a}{x}\right) + (1-x)\log\left(\frac{1-a}{1-x}\right), \text{ where } 0 < a \text{ and } 0 \le x < 1,$$

increases for all $x \le a$.

Using the expression given in Theorem 1 for the coordinates of $\hat{\vec{p}}^{(i_1)}$, consider

$$\log L(\hat{\vec{p}}, \hat{\vec{p}}^{(i_1)}) - \log L(\hat{\vec{p}}, \hat{\vec{p}}) = \sum_{\nu=1}^{i_1} \hat{p}_\nu \log\left(\frac{P_0{}^{i_1}}{\hat{P}^{i_1}}\right) + \sum_{\nu=i_1+1}^{K} \hat{p}_\nu \log\left(\frac{1-P_0{}^{i_1}}{1-\hat{P}^{i_1}}\right) +$$

$$(1-\hat{P}^K)\log\left(\frac{1-P_0{}^{i_1}}{1-\hat{P}^{i_1}}\right) =$$

$$\hat{P}^{i_1}\log\left(\frac{P_0{}^{i_1}}{\hat{P}^{i_1}}\right) + (1-\hat{P}^{i_1})\log\left(\frac{1-P_0{}^{i_1}}{1-\hat{P}^{i_1}}\right) = g_{P_0{}^{i_1}}(\hat{P}^{i_1}).$$

Lemma 3B says that $g_{P_0{}^{i_1}}(\hat{P}^{i_1})$ is an increasing function of $\hat{P}^{i_1}$ when $\hat{P}^{i_1} \le P_0{}^{i_1}$. We keep this in mind in the proof of the next lemma.

LEMMA 4B. *If $\lambda(N)$ is the test statistic for the GLRT discussed here whose acceptance region contains Ω_0, then its rejection region contains the rejection region of a Union-Intersection Test.*

PROOF: Let

$$\mathcal{R}_\delta = \{\log(\lambda(N)) \le \log(\delta)\},$$

the rejection region for the GLRT for some error rate α. Assume that δ and α are such that Ω_0 is contained in $\mathcal{R}_\delta^c$, the acceptance region. Assume that $\hat{p} \in \Omega_0^c$. As defined in the main body of this paper, $\hat{\vec{p}}_0$ satisfies

$$\sup_{\vec{p}\in\Omega_0} \log L(\hat{\vec{p}}, \vec{p}) = \log L(\hat{\vec{p}}, \hat{\vec{p}}_0)$$

and define

$$\text{Region}(i_1, \cdots, i_t) = \{\hat{\vec{p}} \in \Omega_0^c \mid \hat{\vec{p}}_0 = \hat{\vec{p}}^{(i_1,\cdots,i_t)}\}.$$

By definition, $\hat{\vec{p}} \in \text{Region}(i_1, \cdots, i_t) \Rightarrow \hat{\vec{p}} \in \Omega_0^c$. Since

$$\text{Plane}_{\vec{p}_0}^{(i_1,\cdots,i_t)} \subset \text{Plane}_{\vec{p}_0}^{(i_t)},$$

we have

$$\log L(\hat{\vec{p}}, \hat{\vec{p}}^{(i_1,\cdots,i_t)}) \le \log L(\hat{\vec{p}}, \hat{\vec{p}}^{(i_t)}).$$

For such $\hat{\vec{p}}$, the denominator of $\lambda(N)$ is $L(\hat{\vec{p}}, \hat{\vec{p}})$. Subtracting $L(\hat{\vec{p}}, \hat{\vec{p}})$ from both sides of the last inequality yields

$$\log(\lambda(N)) = N\left(\log\mathrm{L}(\hat{\vec{p}}, \hat{\vec{p}}^{(i_1,\cdots,i_t)}) - \log\mathrm{L}(\hat{\vec{p}}, \hat{\vec{p}})\right) \leq N\left(\log\mathrm{L}(\hat{\vec{p}}, \hat{\vec{p}}^{(i_t)}) - \log\mathrm{L}(\hat{\vec{p}}, \hat{\vec{p}})\right) = N\, g_{P_0{}^{i_t}}(\hat{P}^{i_t}). \qquad (A5)$$

From Lemma 2B, for such $\hat{\vec{p}}$, $\hat{P}^{i_t} < P_0{}^{i_t}$ and from Lemma 3B and the discussion that followed, there is a $p^*_{i_t}$, defined by

$$p^*_{i_t} = \max_{0 \leq p \leq P_0{}^{i_t}} \left\{ p \mid g_{P_0{}^{i_t}}(p) \leq \frac{\log(\delta)}{N} \right\}, \qquad (A6)$$

such that

$$g_{P_0{}^{i_t}}(\hat{P}^{i_t}) < \frac{\log(\delta)}{N} \iff \hat{P}^{i_t} < p^*_{i_t}.$$

From (A5), it follows that

$$\hat{\vec{p}} \in \mathrm{Region}(i_1, \cdots, i_t) \text{ and } \hat{P}^{i_t} < p^*_{i_t} \Rightarrow \log(\lambda(N)) < \log(\delta). \qquad (A7)$$

We conclude that

$$\mathrm{Region}(i_1, \cdots, i_t) \cap \{\vec{p} \mid \hat{P}^{i_t} < p^*_{i_t}\} \subset \mathcal{R}_\delta.$$

In the case that $p^*_{i_t} = 0$,

$$Region(i_1, \cdots, i_t) \cap \left\{\hat{p} \middle| \hat{P}^{i_t} < p^*_{i_t}\right\} = \emptyset \subset \mathcal{R}_\delta.$$

By complementation and the definition of $p^*_{i_t}$, it follows that if

$$\hat{\vec{p}} \in \mathrm{Region}(i_1, \cdots, i_t) \cap \mathcal{R}^c_\delta,$$

then

$$\log(\lambda(N)) \geq \log(\delta) \Rightarrow g_{P_0{}^{i_t}}(\hat{P}^{i_t}) \geq \frac{\log(\delta)}{N} \iff \hat{P}^{i_t} \geq p^*_{i_t}.$$

If there exists a subset of $\{1, 2, \cdots, K\}$ containing i_t such that

$\mathrm{Region}(i_1, \cdots, i_t) \cap \mathcal{R}^c_\delta \neq \emptyset$, for such $\hat{\vec{p}}$ we have the inequality $p^*_{i_t} < \hat{P}^{i_t} < P_0{}^{i_t}$

and conclude that the acceptance region of $T(p^*_{i_t})$ contains the set $\{\vec{p} \mid P^{i_t} \geq P_0{}^{i_t}\}$;

$$\{\vec{p} \mid P^{i_t} \geq P_0{}^{i_t}\} \subseteq \{\vec{p} \mid \hat{P}^{i} \geq p^*_{i_t}\}.$$

Define S_i = collection of all subsets of the set $\{1, 2, \cdots, K\}$ that contain i. Note that

$$\Omega^c_0 = \mathrm{Region}(1) \cup \mathrm{Region}(2) \cup \mathrm{Region}(1,2) \cup \mathrm{Region}(3) \cup$$

$$\text{Region}(1,3) \cup \text{Region}(2,3) \cup \cdots \cup \text{Region}(1,2,\cdots,K).$$

Equivalently, we write

$$\Omega_0^c = \bigcup_{i=1}^{K} \bigcup_{S \in S_i} \text{Region}(S).$$

If for each i, $1 \leq i \leq K$, $\text{Region}(S) \neq \emptyset$ for some $S \in S_i$, we have

$$\mathcal{U}_\delta^c \equiv \bigcap_{i=1}^{K} \{\hat{\vec{p}} \mid \hat{P}^i \geq p_i^*\} \supseteq \bigcap_{i=1}^{K} \{\hat{\vec{p}} \mid \hat{P}^i \geq {P_0}^i\} = \Omega_0. \tag{A8}$$

Note that $\mathcal{U}_\delta^c$ is the acceptance region of a union-intersection test characterized by the collection of tests

$$\{T(p_i^*),\ 1 \leq i \leq K\}.$$

It follows from (A8) that

$$\mathcal{U}_\delta \subset \Omega_0^c = \bigcup_{i=1}^{K} \bigcup_{S \in S_i} \text{Region}(S).$$

If $\hat{p} \in \mathcal{U}_\delta$ then $\hat{p} \in S$ for $S \in S_i$, for some i, $1 \leq i \leq K$. From (A8), we deduce that

$$\hat{P}^i < p_i^* \Rightarrow \log\big(\lambda(N)\big) < \log(\delta) \Rightarrow \hat{p} \in \mathcal{R}_\delta.$$

We conclude that the rejection region of this union-intersection test is contained in the rejection region of the GLRT or the acceptance region of the GLRT is contained in the acceptance region of this union-intersection test.

It has been assumed that for each $i, 1 \leq i \leq K$, at least one of the sets in S_i is nonempty. This is needed to show (A8). If there exists an i such that S_i is a collection of empty sets, then for no $\hat{\vec{p}} \in \Omega_0^c$ is the likelihood function maximized on Edge(i) or on any of Edge(i)'s intersections with other edges. The assumption ensures that the acceptance region of the GLRT contains Ω_0. For such an i, we let $p_i^* = {P_0}^i$,. then (A8) continues to be true, ensuring the above-stated relationship between Ω_0 and $\mathcal{U}_\delta^c$. ◇

The test that accepts H_0 iff $\hat{\vec{p}} \in \Omega_0$ is also a union-intersection test. In this case $p_i = {P_0}^i$, $1 \leq i \leq K$. We have have proven the following theorem:

<u>THEOREM 3</u>. *If $\lambda(N)$ is the test statistic for the GLRT, the acceptance region of which contains Ω_0, then its rejection region, $\mathcal{R}_\delta$, is bounded respectively above and below in terms of set inclusion by the rejection regions of Union- Intersection Tests*

$$\bigcap_{\nu=1}^{K} T(p_\nu^*) \quad \text{and} \quad \bigcap_{\nu=1}^{K} T({P_0}^\nu).$$

◇